Clinical Psychology Review 83 (2021) 101944

Contents lists available at ScienceDirect

Clinical Psychology Review

journal homepage: www.elsevier.com/locate/clinpsychrev

Review

Are videoconferenced mental and behavioral health services just as good as in-person? A meta-analysis of a fast-growing practice

Ashley B. Batastini[a,b,*,1], Peter Paprzycki[a,c,d], Ashley C.T. Jones[a], Nina MacLean[e]

[a] University of Southern Mississippi, USA
[b] University of Mississippi Medical Center Department of Psychiatry, USA
[c] Mississippi Center for Clinical and Translational Research, USA
[d] University of Toledo, USA
[e] Michigan Department of Health & Human Services - Center for Forensic Psychiatry, USA

ARTICLE INFO

Keywords:
Telepsychology
Telepsychiatry
Telemental health
Videoconferencing
Meta-analysis

ABSTRACT

The use of videoconferencing technologies (VCT) is on the rise given its potential to close the gap between mental health care need and availability. Yet, little is known about the effectiveness of these services compared to those delivered in-person. A series of meta-analyses were conducted using 57 empirical studies (43 examining intervention outcomes; 14 examining assessment reliability) published over the past two decades that included a variety of populations and clinical settings. Using conventional and HLM3 meta-analytical approaches, VCT consistently produced treatment effects that were largely equivalent to in-person delivered interventions across 281 individual outcomes and 4336 clients, with female clients and those treated in medical facilities tending to respond more favorably to VCT than in-person. Results of an HLM3 model suggested assessments conducted using VCT did not appear to lead to differential decisions compared to those conducted in-person across 83 individual outcomes and 332 clients/examinees. Although aggregate findings support the use of VCT as a viable alternative to in-person service delivery of mental healthcare, several limitations in the current literature base were revealed. Most concerning was the relatively limited number of randomized controlled trials and the inconsistent (and often incomplete) reporting of methodological features and results. Recommendations for reporting the findings of telemental health research are provided.

1. Introduction

Although various forms of remote and mobile services have begun to infiltrate the practice of psychology and psychiatry, the use of videoconferencing technology (VCT) has increased rapidly over the past decade (and exponentially in the past few months alone), with trends in the use of these systems expected to continue well into the new decade (Norcross, Pfund, & Prochaska, 2013). Specifically, VCT uses "real-time" audiovisual monitors/screens to connect agencies or clients in need of services to providers who can render such services (Ax et al., 2007). In fact, the November 2017 issue of *The Monitor on Psychology* listed the integration of technology into psychological practice as a top 10 trend in the field. Others have hailed remote technologies including VCT as the "key to solving mental healthcare access problems in the twenty-first century" (Frueh, 2015, p. 304) and a "modern answer to mental health" (Matthews, 2017).

The use of remote healthcare proliferated over a relatively brief period of time. For example, in 1991, there were only four telemedicine networks across the United States; merely five years later, there were approximately 160 (Miller, Clark, Veltkamp, Burton, & Swope, 2008). Among psychologists, the use of VCT increased from 2% in 2000 to 10% in 2008 (APA Psychology Health Service Provider Survey, 2008). In a more recent survey, nearly 40% of behavioral health providers in the U. S. reported using VCT as adjunctive to in-person services, and almost 45% had used VCT independently (Gershkovich, Herbert, Forman, & Glassman, 2016). The use of such technology is a popular response to rising health care costs and the need to increase access to care with qualified professionals and specialists, especially in rural communities (Nordal, 2015). While much of the research to date is based on U.S. practices and clientele, the fast-growing use of VCT can be observed

* Corresponding author at : Counseling, Educational Psychology & Research, 100 Ball Hall, Memphis, TN 38152, USA.
E-mail address: Ashley.batastini@memphis.edu (A.B. Batastini).

[1] Dr. Batastini was primarily affiliated with the University of Southern Mississippi at the time this study was completed

https://doi.org/10.1016/j.cpr.2020.101944
Received 30 March 2020; Received in revised form 19 October 2020; Accepted 2 November 2020
Available online 17 November 2020
0272-7358/

internationally (see e.g., De Las Cuevas, Arredondo, Cabrera, Sulzenbacher, & Meise, 2006; Modai et al., 2006; Yoon et al., 2018). In particular, Australia (Australia's National Digital Health Strategy, 2018) and the United Kingdom (Digital Health & Care Scotland, 2018; United Kingdom National Health Service, 2019) have also been at the forefront of virtual mental health technologies. In Canada, one study estimated reductions associated with telepsychiatric services at an average of $50 per visit (O'Reilly et al., 2007).

In the U.S., the high demand for mental health services and increased acceptance of virtual modalities has resulted in legislative efforts to make it easier for telehealth technologies to reach across state lines. Established in 2014, the Interstate Medical Licensure Compact (IMLC) allows qualifying physicians to practice remotely in up to 29 states that are part of the agreement without obtaining licensure in those states. Following suit, the Psychology Interjurisdictional Compact (PSYPACT; Lerman, Kim, Ozinal, & Thompson, 2018) was introduced to similarly allow licensed psychologists in participating states to provide remote services for specific and limited purposes in other participating states. At the time of this publication, 12 states had signed onto PSYPACT and 12 others had pending PSYPACT legislation. Regardless, nearly all state professional boards have set some parameters for inter-jurisdictional remote mental and behavioral health practice (Lerman et al., 2018). In keeping up with the surgency of remote services, various medical and mental health organizations, including the American Psychological Association (2013), the American Telemedicine Association (2013), and the American Psychiatric Association (2018) have also established guidelines for the ethical and secure practice of mental health delivered over a distance. Given the current global health crisis, it would be remiss not to acknowledge the near overnight shift in practitioners turning to virtual services, frantically needing to acquire adequate training and infrastructure to connect with clients. Accompanying this shift have been expansions to Medicare coverage such that, for the duration of the COVID-19 emergency, real-time telehealth visits can be billed at the same rate as in-person visits (Centers for Medicare and Medicaid Services, 2020), as well as the passing of a congressional bill to establish home-based telemental health care (S. 3917, 2020) that make mental health services more accessible. Similar changes were also observed globally as other countries such as China (Liu et al., 2020), Australia (Zhou et al., 2020), and France (Ohannessian, Duong, & Odone, 2020) sought to expand and encourage the use of telemedicine during the COVID-19 emergency. This pandemic is likely to forever alter the way clinicians view and deliver mental healthcare.

VCT is not a new technology. In fact, elements of VCT were first introduced in medical settings across the U.S. for the purpose of diagnosis, patient care, and training over 30 years ago (Miller et al., 2008). As technology has progressed, an increasing number of studies have examined the efficacy of this modality compared to in-person intervention and assessment of mental and behavioral health concerns. The majority of existing studies have examined the success of VCT in the treatment of specific disorders, such as anxiety disorders, substance use, depression, and eating disorders (e.g., Benavides-Vaello, Strode, & Sheeran, 2013; Osenbach, O'Brien, Mishkind, & Smolenski, 2013; Rees & Maclaine, 2015; Sproch & Anderson, 2019). Other studies have focused on using VCT with special populations for whom the gap between service need and availability is especially wide or for specific types of services (e.g., neuropsychological assessment; see Brearly et al., 2017). The most common population for which VCT has been studied appears to be veterans (Gros, Yoder, Tuerk, Lozano, & Acierno, 2011; Luxton, Nelson, & Maheu, 2016), though several studies have also focused on forensic clients and inmates (see Batastini, McDonald, & Morgan, 2013 for a review). In general, current research on VCT-delivered mental and behavioral health services has produced promising results, supporting the idea that these services are largely as effective as in-person. These findings appear to be good news considering the growing expectation that remote services, and VCT in particular, will help combat the mental health care crisis that is projected to cost the U.S. nearly $750 billion dollars in 2021 by increasing access to affordable psychological and psychiatric care, reducing wait times, and offering better continuity of care (LaRock, 2019).

Despite the growing use of and support for VCT, along with other forms of remote mental health care (e.g., online support groups, mobile apps), there remains comparatively little empirical evidence on its effectiveness. Although most studies that examine video-based services are promising, the aggregate effects are not fully understood. To date, multiple meta-analyses have examined the overall efficacy of telemental health within specific populations and diagnostic presentations (e.g., Batastini & Morgan, 2016; Brearly et al., 2017; Larson, Rosen, & Wilson, 2018; Osenbach et al., 2013; Sloan, Gallagher, Feinstein, Lee, & Pruneau, 2011). Another looked specifically at objective psychiatric assessments, finding no differences in accuracy or satisfaction across modalities (Hyler, Gangure, & Batchelder, 2005). In fact, satisfaction and therapeutic alliance have been broadly examined in various countries, including Canada (Germain, Marchand, Bouchard, Guay, & Drouin, 2010), England (Manchanda & McLaren, 1998), Australia (Stubbings, 2012) and Scotland (Simpson, Bell, Knox, & Mitchell, 2005). The only known meta-analysis comparing therapy and assessment outcomes across VCT and in-person providers (Drago, Winding, & Antypa, 2016) focused strictly on psychiatric services. Importantly, given differing educational and training models of each discipline, psychological and psychiatric services can differ quite dramatically with regard to therapeutic orientation, the client-doctor relationship, duration of services, and structure of services, among other variables. Therefore, meta-analyses that do not include literature related to psychological approaches are limited in scope and do not fully capture the range or intensity of services that are now routinely provided through VCT. Other systematic reviews of the effectiveness of telemental health also exist that attempt to take a broader look at the practice (e.g., Hilty et al., 2013; Langarizadeh et al., 2017; Rees & Maclaine, 2015; Salmoiraghi & Hussain, 2015); however, these only provide descriptive statistics and more of a narrative overview rather than a statistically controlled comparison between traditional in-person and VCT services.

The current meta-analysis adds a more generalist perspective to the literature base that has not yet been captured empirically. That is, our primary aim was to answer the basic question of whether VCT is better, worse, or relatively equivalent to in-person mental health services. Taking a more comprehensive, aggregate approach is needed for several reasons. First, it allows for an additive, bigger picture interpretation; can we say VCT is equally effective in an overall sense *and* for specific clinical purposes or clients? Second, knowing the overall effects of VCT can serve as point of comparison to better contextualize effect sizes produced by narrowly focused meta-analyses; that is, how do aggregate effects for specific clinical purposes or clients measure up to those for VCT in general? Third, working with a larger collection of studies offers an opportunity for more robust comparisons of effects across different variables of interest than what can be performed within smaller-scale meta-analyses; for example, are outcomes for certain disorder categories stronger or weaker compared to others? The ability to examine these comparisons likewise provides a meaningful context within which to interpret findings from other analyses of specialized client groups. Fourth, and related to comparability, because this meta-analysis captures studies published before the dramatic shifts associated with COVID-19, it can serve as a historical baseline for future meta-analyses that are sure to follow from the inevitable spike in empirical studies that will emerge during and post-COVID-19. Finally, we view the present study as a wholistic effort to reveal what is out there and to what extent and, conversely, uncover the types of clinical services and populations that are less represented in the literature. While extant systematic reviews can help with this delineation, they lack accompanying statistical insights about the actual potential of VCT to engender relatively equivalent (or distinct) outcomes to in-person.

Several factors contribute to the uniqueness of the present set of meta-analyses. First, we applied comprehensive inclusionary criteria such that studies addressing a wide range of mental health disorders, client populations, interventions, and assessment types are represented. Second, we were concerned only with studies that compare VCT services to an in-person control; pre-post studies examining only VCT were excluded to clearly and directly answer the question of whether VCT produces outcomes that are comparable to in-person rather than whether VCT is an improvement over the absence of services. Further, in examining outcomes explicitly associated with therapeutic interventions, we apply multiple contemporary meta-analytic strategies and examine several moderator analyses that further explore what works and for whom. Overall, we hypothesized that meta-analytic results would support VCT as an equally effective treatment and assessment modality compared to in-person services by yielding small and/or non-significant effect size estimates when comparing relevant mental health outcomes by modality type.

Perhaps equally important to the meta-analytic findings themselves, the process of compiling empirical work products uncovered several limitations and flaws in the available literature that we argue must be discussed and addressed in future evaluation efforts. Following the presentation of our primary findings, we emphasize the need for more controlled treatment outcome and assessment reliability studies and offer initial reporting recommendations for future researchers. As will be detailed later, a large number of studies were excluded for lack of a comparison group, because they did not focus directly on service efficacy (e.g., instead asking about service satisfaction or acceptability), or because useful data was unknown or reported in a limited manner. While some guidelines exist regarding the empirical study of telemedicine practices (Krupinski & Bernard, 2014) and there are general manuscript reporting guidelines (e.g., Centers for Medicare and Medicaid Services, 2020), standardized recommendations for reporting findings specifically from telemental health research to scientific consumers, service providers, legislators, and other stakeholders is lacking. One available set of recommendations focuses exclusively on reporting VCT studies for depression interventions (Abel, Glover, Brandt, & Godleski, 2017). While many of these recommendations may apply to this line of research generally, they may also miss important factors or considerations highlighted by work in other domains. Ensuring the systematic and consistent reporting of such findings will not only deepen our understanding of VCT as an alternative service modality but will also increase our confidence in embracing VCT as the new norm in mental healthcare.

2. Method

2.1. Study inclusion

To be included in this meta-analysis, studies had to (1) be published or available in English, (2) evaluate mental health-related outcomes (e.g., symptom reduction, hospitalizations) following a mental health service (i.e., psychiatric consultations, psychotherapy/counseling, assessment), (3) use a telecommunication service delivery system that transmitted live audio and visual information simultaneously, (4) use a between-groups comparison design, and (5) report sufficient information to allow for a calculation of effect size estimates. When data were insufficiently described in a published report, the study's corresponding author was contacted in an attempt to ascertain needed details. Studies with only within-subjects designs (i.e., pre-post telehealth, waitlist control) or that examined a nonmental health service (e.g., physical health services) were excluded from this review.

2.2. Selection strategy

Keywords primarily related to telehealth, telemedicine, telepsychology, telemental health, and telepsychiatry were entered into 38 electronic databases and internet search engines (e.g., PsycINFO, Medline, criminal justice abstracts, Google scholar, PsycCRITIQUES, and Science & Technology Collection). Of note, due to the large number of hits for "telehealth" and "telemedicine" (e.g., at the time of this publication Medline yielded 21,906 and 30,117 results, respectively; Google scholar currently generates over 120,000 and 600,000 results, respectively), these broader terms were paired with more specific search terms (e.g., *telehealth* AND *psych**, *eHealth* AND *depression*). Terms like "telepsychology" and "telemental health" were searched separately. Reference lists of review articles (e.g., Antonacci, Bloch, Saeed, Yildirim, & Talley, 2008; Backhaus et al., 2012; Drago et al., 2016; Osenbach et al., 2013; Sloan et al., 2011; Young, 2012) and chapters in an edited book on telemental health services (Myers & Turvey, 2012) were also examined

An initial search of document titles yielded a total of 504 related articles that were maintained for a further review. All studies were then preliminarily categorized as either "meets criteria," "does not meet criteria," or "unsure" based on their abstracts. An examination of interrater reliability for this stage of the sorting process Using 50 randomly selected articles (about 10%) Yielded an intraclass correlation of 0.87. The first author (ABB) and another doctoral-level researcher (CMK; see author note) then verbally discussed each article categorized as "uncertain" to determine whether it could be clearly eliminated or should be maintained for double coding. All inclusionary/exclusionary decisions that were made prior to double coding took a liberal approach to help ensure articles were not mistakenly excluded. An updated literature search concluded in may 2020 to capture any studies that may have been overlooked or published during the first-round double-coding and reliability process (detailed below). Notably As the collection of studies on which this meta-analysis is based represents a program of research expanding several years Some articles were informally accumulated over time and therefore not obtained from intentional literature searches. An additional 52 articles had been obtained between the first-round search and the updated search ending in early 2020. In total 65 articles (47 examining intervention outcomes; 18 examining assessment reliability) were deemed appropriate for inclusion in the quantitative meta-analyses. During the analysis phase Eight additional articles were found to have insufficient or incomplete statistical data for inclusion in the planned analyses Leaving a final total of 57 studies (43 intervention; 14 assessment)

Studies that met all inclusionary criteria and were used in the final analyses were published between 1997 and 2019, representing over 20 years' worth of research. Two studies published in 2020 were identified (Glynn, Chen, Dawson, Gelman, & Zeliadt, 2020; Halphen et al., 2020) that described the implementation of telehealth programs for pain management and capacity evaluations, respectively. However, neither study reported quantitative outcome data nor conducted comparisons to in-person treatment. Therefore, no studies published between January and March of 2020 met inclusionary criteria. In most cases, studies that were excluded from analyses either lacked adequate methodological features (e.g., did not use an appropriate comparison or control group), measured outcomes in a way that could not be meaningfully combined to calculate effect size estimates, or reported insufficient statistical details (in conjunction with study authors either not returning requests for additional information or no longer having access to the data). The final inclusion rate of 45.6% (57 out of 125) in this study is higher than at least two other meta-analyses of telepsychological services for specific populations (31.7%; Sloan et al., 2011; 9.7%, Osenbach et al., 2013). For simplicity, counts shown in Fig. 1 were collapsed across all stages of article identification and exclusion.

2.3. Data coding and extraction

Coding was completed in two stages. In both stages, articles identified as meeting inclusionary criteria were randomly assigned to two independent coders trained in the use of a standardized coding form. Two randomly selected articles were used in the training process. The training articles also served to pilot the coding form and ensure it was structured to reasonably accommodate variations across studies. After

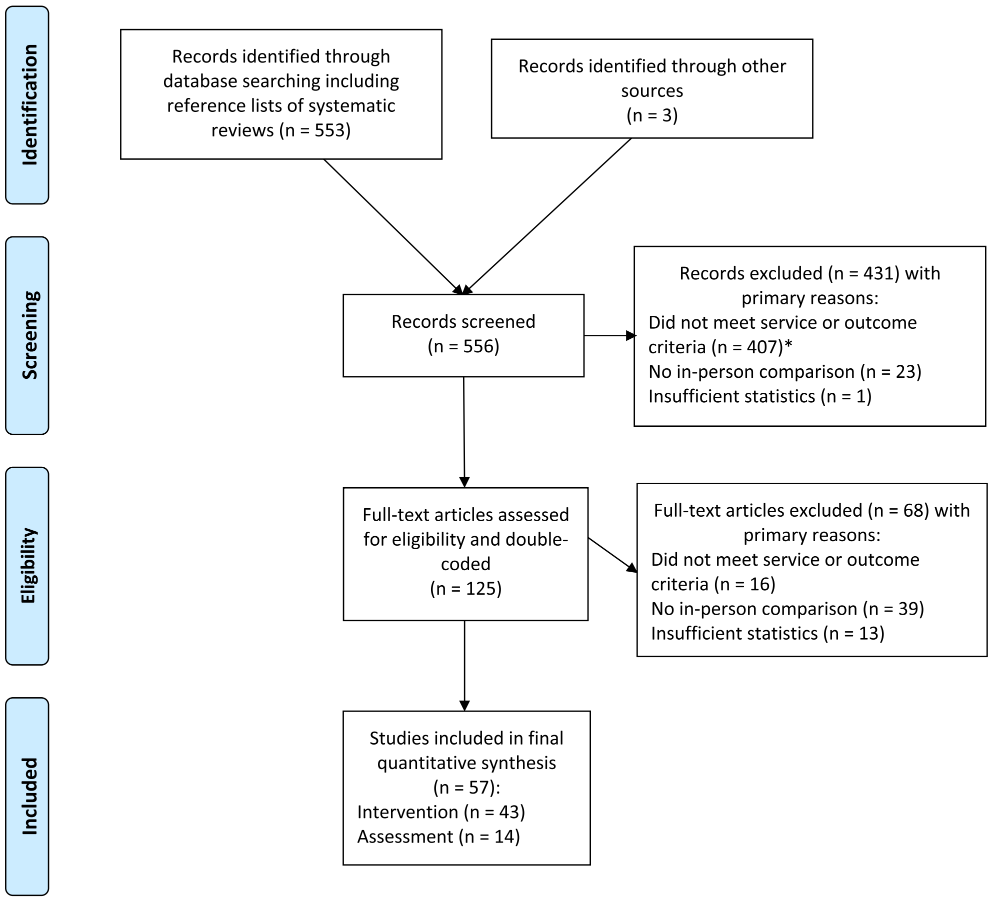

Fig. 1. Flow chart of study inclusion/exclusion decision tree.
*Although some articles could have also been excluded due to the lack of an in-person comparison group or insufficient statistical data, they were coded in this category because information about the service, modality, or type of outcome evaluated was typically the first problem identified (based on the title or abstract). Once ineligibility was established, no additional reasons were recorded.

independent codes were submitted, discrepancies between rows of data for each of the two assigned coders were compared. Discrepancies were highlighted in an Excel spreadsheet (by a separate trained research assistant) and discussed in-person between the coders to determine the best possible response. When coders were unable to agree during these verbal discussions, the first author (ABB) was consulted.

The first stage of data extraction involved coding general study descriptive information and removing any articles that failed to meet inclusionary criteria but were not identified as such during screening procedures. Although many variables ultimately were not consistently reported within articles, we attempted to code a wide range of descriptive information about each study including general reference information (e.g., article title, authors, publication year, population of interest); site descriptors (e.g., where services were provided vs. where they were received); technology used (e.g., software program, data transmission network, quality of video resolution); sample descriptors (e.g., sample size, attrition rate, demographic and clinical composition of sample); therapist/evaluator descriptors (e.g., gender, ethnicity, educational level); treatment descriptors (e.g., type of service, number of sessions, length of sessions); and research design (e.g., mixed- or between-subjects, method of group assignment).

The second coding stage involved extracting specific outcomes of interest within each study related to the effectiveness of the target intervention or assessment reliability. Outcomes of interest were coded by type and corresponding test statistics (means, standard deviations, measures of effect size). The majority of treatment-related outcomes were classified as self-reported mental health symptoms (i.e., scores from a measure of mental health completed by the client/patient; e.g., the Beck Depression Inventory, Brief Symptom Inventory, PSTD Checklist). Other outcomes included provider observations of a client's behavior or functioning (e.g., GAF score, Mini Mental Status Exam), medication or treatment compliance (e.g., substance use relapse), and frequency of psychiatric hospitalizations or office visits (see Table 1). For assessment reliability articles, raw data (means, standard deviations, frequencies) that were used to calculate reliability coefficients across assessment modalities were extracted. Only advanced doctoral-level coders, and the first (ABB) and second (PP) authors participated during this stage given their more extensive knowledge of statistical analyses and concepts. All statistical outcomes were coded in such a way that a positive effect size always favored the in-person group and negative effect sizes favored the VCT group. Thus, for the intervention-based studies, negative values indicate greater pre-post intervention gains of the VCT intervention group compared to the in-person intervention group—however, most outcome measures (regardless of treatment modality) were expected to be reduced by post-assessment in comparison to baseline (e.g., a reduction in anxiety symptoms was expected following exposure to either intervention). For the assessment reliability studies, negative values indicate that, when clients were assessed via VCT, they were evaluated as less symptomatic or healthier than when they were assessed in-person. Similarly, however, it was expected that any discrepancies in evaluation outcomes across modality would be minimal and non-significant.

3. Statistical strategy

3.1. Intervention studies

All intervention outcome measures (for VCT and in-person) were converted to Hedges' *g* standardized effect sizes. Most of the intervention outcomes ($n = 256$ or 91.1%) were reported as means and corresponding standard deviations (*SD*s) or standard errors (*SE*s). These outcomes were converted to Hedges' *g* indices using the pooled *SD*s of

Table 1
Methodological features of all meta-analyzed studies.

Study	Type of outcome(s)	Participant type	Study design	Software used	Type of remote site	Outcome(s) of interest category
Acierno et al. (2016)	Treatment outcomes	Clients	Between-subjects	NA	Home or university clinic	Self-reported mental health symptoms
Batastini and Morgan (2016)	Treatment outcomes	Other: inmate clients	Mixed	Polycom	Correctional facility/ prison/jail (including internal psychiatric unit)	Self-reported mental health symptoms
Bouchard et al. (2004)	Treatment outcomes	Clients	Mixed	Cisco/Tanberg	Outpatient	Self-reported mental health symptoms; Self-reported medical/physical symptoms
Chapman et al. (2019)	Assessment	Mixed	Within-subjects	Zoom	Home or university clinic	Provider observations of mental health symptoms
Choi et al. (2014)	Treatment outcomes	Clients	Mixed	Skype	Home or university clinic	Self-reported mental health symptoms
Chong & Moreno et al. (2012)	Treatment outcomes	Clients	Between-subjects	Other: Macromedia Breeze Manager	Medical facility	Self-reported mental health symptoms
Comer et al. (2017)	Treatment outcomes	Mixed	Mixed	NA	NA	Self-reported mental health symptoms; provider observations of mental health symptoms; caregiver observations of mental health symptoms
Comer et al. (2017)	Treatment outcomes	Mixed	Mixed	NA	Home or university clinic	Self-reported mental health symptoms; provider observations of mental health symptoms; caregiver observations of mental health symptoms
Day and Schneider (2002)	Treatment outcomes	Clients	Between-subjects	NA	Home or university clinic	Self-reported mental health symptoms; Provider observed mental health symptoms
De la Cuevas, Arrendondo, Cabrera, Sulzenbacher, and Meise (2006)	Treatment outcomes	Clients	Mixed	Polycom	Outpatient	Self-reported mental health symptoms
Dekhtyar, Braun, Billot, Foo, and Kiran (2019)	Reliability	Mixed	Within-subjects	Other: GoToMeeting, Zoom	Home or university clinic	Provider observations of mental health symptoms
Elford et al. (2000)	Reliability	Mixed	Mixed	Other: PCPoint	Medical facility	Provider observation of mental health symptoms; Provider opinion of treatment needs
Farabee, Calhoun, and Veliz (2016)	Treatment outcomes	Clients	Mixed	NA	Outpatient	Self-reported mental health symptoms
Fortney et al. (2007)	Treatment outcomes	Clients	Between-subjects	NA	Veteran hospital	Medication/treatment compliance/non-compliance
Fortney, Maciejewski, Tripathi, Deen, and Pyne (2011)	Treatment outcomes	Clients	Between-subjects	NA	Veteran hospital	Frequency of psychiatric hospitalization; Frequency of doctor visits
Fortney et al. (2013)	Treatment outcomes	Clients	Mixed	NA	Medical facility	Medication/treatment compliance/non-compliance; Frequency of doctor visits
Frueh et al. (2007)	Treatment outcomes	Clients	Mixed	Polycom	Veteran hospital	Provider observed mental health symptoms; Medication/treatment compliance/non-compliance
Germain, Marchand, Bouchard, Drouin, and Guay (2009)	Treatment outcomes	Clients	Mixed	Cisco/Tanberg	NA	Provider observed mental health symptoms
Glassman et al. (2019)	Treatment outcomes	Clients	Mixed	Cisco/Tanberg	Veteran hospital	Self-reported mental health symptoms
Herbert et al. (2017)	Treatment outcomes	Clients	Mixed	NA	Veteran hospital	Self-reported mental health symptoms
Himle et al. (2012)	Treatment outcomes	Mixed	Mixed	Sony	Home or university clinic	Provider observations of mental health symptoms; caregiver observations of mental health symptoms
Hulsbosch, Nugter, Tamis, and Kroon (2017)	Treatment outcomes	Clients	Mixed	NA	Home or university clinic	Self-reported mental health symptoms
Hungerbuehler, Valiengo, Loch, Rössler, and Gattaz (2016)	Treatment outcomes	Clients	Mixed	Skype	Home or university clinic	Self-reported mental health symptoms
Kelleher et al. (2019)	Treatment outcomes	Clients	Mixed	NA	NA	Self-reported mental health symptoms
King, Brooner, Peirce, Kolodner, and Kidorf (2014)	Treatment outcomes	Clients	Mixed	eGetgoing	Outpatient	Documented substance use/relapse
Kobak, Williams, and Engelhardt (2008)	Reliability	Clients	Within-subjects	Polycom	NA	Self-reported mental health symptoms; Provider observed mental health symptoms
Liu et al. (2019)	Treatment outcomes	Clients	Mixed	NA	Outpatient	Self-reported mental health symptoms; self-reported medical/physical symptoms
Luxton et al. (2016)		Clients		Cisco/Tanberg		Self-reported mental health symptoms

(*continued on next page*)

Table 1 (*continued*)

Study	Type of outcome(s)	Participant type	Study design	Software used	Type of remote site	Outcome(s) of interest category
	Treatment outcomes		Between-subjects		Home or university clinic	
Maieritsch et al. (2016)	Treatment outcomes	Clients	Mixed	NA	NA	Provider observations of mental health symptoms; Self-reported mental health symptoms
Mitchell et al. (2008)	Treatment outcomes	Clients	Mixed	NA	NA	Self-reported mental health symptoms; Self-reported medical/physical symptoms
Modai et al. (2006)	Treatment outcomes	Other: Physicians	Mixed	Other: Falcon Internet protocol version 3	Medical facility	Self-reported mental health symptoms; Provider observed mental health symptoms
Montani et al. (1997)	Reliability	Mixed	Within-subjects	NA	Medical facility	Provider observations of mental health symptoms
Moreno et al. (2012)	Treatment outcomes	Clients	Mixed	Other: fax and standard computer setup	Medical facility	Self-reported mental health symptoms
Morgan, Patrick, and Magaletta (2008)	Treatment outcomes	Mixed	Between-subjects	Cisco/Tanberg	Correctional facility/ prison/jail (including internal psychiatric unit)	Self-reported mental health symptoms; Medication/treatment compliance/non-compliance
Morland et al. (2010)	Treatment outcomes	Clients	Mixed	NA	Medical facility	Self-reported mental health symptoms; Provider observed mental health symptoms
Morland et al. (2014)	Treatment	Clients	Mixed	Cisco/Tanberg	Veteran hospital	Self-reported mental health symptoms; treatment compliance/non-compliance
Morland et al. (2015)	Treatment outcomes	Clients	Between-subjects	Cisco/Tanberg	Veteran hospital	Self-Reported Mental Health Symptoms
Nelson, Barnard, and Cain (2003)	Treatment outcomes	Clients	Between-subjects	PictureTel	Home or university clinic	Self-reported mental health symptoms, Suicidal ideation/attempts
O'Reilly et al. (2007)	Treatment outcomes	Clients	Between-subjects	Polycom	Medical facility	Self-reported mental health symptoms; Frequency of psychiatric hospitalization
Porcari et al. (2009)	Reliability	Clients	Between-subjects	Cisco/Tanberg	Veteran hospital	Provider observation of mental health symptoms
Ruskin et al. (2004)	Treatment	Clients	Between-subjects	Other: VTEL	Veteran hospital	Self-reported mental health symptoms
Shulman, John, and Kane (2017)	Treatment outcomes	Clients	Mixed	Other: Webex	Home or university clinic	Medication/treatment compliance
Simpson, Guerrini, and Rochford (2015)	Treatment outcomes	Clients	Mixed	Cisco/Tanberg	Outpatient	Self-reported mental health symptoms
Stain et al. (2011)	Reliability	Clients	Mixed	NA	NA	Provider observation of mental health symptoms
Stead and Vinson (2019)	Reliability	Mixed	Within-subjects	Other: Facetime	Home or university clinic	Provider observation of mental health symptoms
Stubbings, Rees, Roberts, and Kane (2013)	Treatment outcomes	Clients	Mixed	Other: iChat	Home or university clinic	Self-reported mental health symptoms
Temple, Drummond, Valiquette, and Jozsvai (2010)	Reliability	Clients	Within-subjects	Polycom	Other	Provider observation of mental health symptoms
Tse, Mccarty, Stoep, and Myers (2015)	Treatment outcomes	Clients	Mixed	NA	Medical facility	Self-reported mental health symptoms
Tuerk, Yoder, Ruggiero, Gros, and Acierno (2010)	Treatment outcomes	Clients	Between-subjects	Cisco/Tanberg	Veteran hospital	Self-reported mental health symptoms
Vahia et al. (2015)	Reliability	Mixed	Mixed	NA	Medical facility	Provider observations of mental health symptoms
Wadsworth et al. (2016)	Reliability	Clients	Between-subjects	Polycom	Medical facility	Provider observations of mental health symptoms
Wierwille, Pukay-Martin, Chard, and Klump (2016)	Treatment outcomes	Clients	Between-subjects	NA	Veteran hospital	Provider observations of mental health symptoms; Self-reported mental health symptoms.
Wong, Martin-Khan, Rowland, Varghese, and Gray (2012)	Reliability	Clients	Between-subjects	Sony	Other	Provider observations of mental health symptoms
Yoshino et al. (2001)	Reliability	Clients	Between-subjects	Other: Microsoft Netmeeting 2.1	NA	Provider observations of mental health symptoms
Yuen et al. (2015)	Treatment outcomes	Clients	Mixed	Skype	Home or university clinic	Self-reported mental health symptoms
Zarate et al. (1997)	Reliability	Mixed	Between-subjects	Picture Tel	Inpatient psychiatric facility	Provider observations of mental health symptoms

the pre-measures, when available. The method of standardization based on pre-*SDs* is preferable to one based on the post-*SDs* measures because of a possible subject by treatment interaction, which makes the post-*SDs* typically too large (Hunter & Schmidt, 2015). In rare instances, where *SDs* for pre-measures were not available, post-*SDs* were used for standardization. The remaining portion of outcomes ($n = 25$ or 8.9%) were reported as counts of events and non-events (e.g., numbers of patients with depressive symptoms in pre- and post-intervention types, etc.). These outcomes were converted into effect size estimates using ratios of logged odds or natural logarithm of the numerator—i.e., the odds (or the ratio of events to non-events) in the VCT group—and denominator—i.e., the odds (or the ratio of events to non-events) in the in-person group. All effect size computations were carried out manually in Excel and verified for accuracy using the input worksheets of the Comprehensive Meta-Analysis (CMA) software (Borenstein, Hedges, Higgins, & Rothstein, 2014).

A nested structure of participants within stochastically-dependent outcomes and outcome measures within studies (assumed to be independent) called for a multi-level (i.e., sampling-, outcome- and study-level) statistical approach to data analysis. However, prior to conducting a multi-level analysis, a conventional meta-analysis was performed on aggregated outcomes within studies with manual adjustments to effect sizes and their corresponding variances using a procedure proposed by Hedges and Olkin (1985).[2] This approach was taken first due to the availability of more robust methods of assessing outliers and case diagnostics as well as publication bias analysis in the conventional approach than what is currently afforded in the multi-level analysis approach.

The necessary computations involving matrix algebra were conducted in the R program (R Core Team, 2019). The effect size variances of outcomes nested in each study were set to the average value of variances of all outcomes for that given study. Using the average was considered a safe assumption, as the differences in individual effect size variances nested within a particular study were negligible because an approximately equal number of participants in either VCT or in-person conditions ($n_{\text{treatment}}$ and n_{control}, respectively) was observed (e.g., if repeated outcomes were assessed over time, small attrition of participants was observed). To account for correlations between measures, the correlation coefficient for all off-diagonal elements was set to 0.5, indicating a moderate amount of dependence in the outcomes. This standard was selected based on existing research examining correlations between psychological constructs commonly found in the studies used in this meta-analysis (see e.g., Blevins, Weathers, Davis, Witte, & Domino, 2015; Bovin et al., 2016). For instance, correlations between the PCL-5 and other psychological measures such as the BDI, BAI, and AUDIT (self-report measures used in meta-analyzed studies), ranged from $r = 0.25$ to 0.60 (Wortmann et al., 2016). Setting the correlation diagonals in the mid-range was also intended to offset correlations that might be stronger or weaker than this estimate. The conventional analyses were analyzed with R *metafor* package, (Viechtbauer, 2010) using the Restricted Maximum-Likelihood (REML) estimator (package default setting).

Publication bias was examined with funnel plots (Sutton, 2009), with observed effect sizes (with and without moderators described below) referenced on the horizontal axis and their corresponding precisions of estimates (standard errors *SE*s and their variances as well as the inverse of *SE*s and their variances) on the vertical axis in the funnel plots. The degree of variability for smaller sample sizes (indicating less precision) was assessed with 95% pseudo confidence intervals (Viechtbauer, 2010).

Plots of standardized residuals (the standardized difference between the average effect size and the *i*th study's effect size with residual values larger than ±1.96 generally considered as influential), DFFITS (the predicted average effect size change in *SD*s unit as a function of removing an *i*th study with values over 1 generally signifying an influential case), Cook's values (an index that combines the effect of the removal of an *i*th study on leverage and fit) were graphically examined. In addition, covariance ratios (the predicted average change in variance-covariance matrix of estimated parameters with values less than 1 indicative of cases that reduce the precision of parameters estimates), variance or τ^2 (the predicted changes in the estimate of total heterogeneity with the removal of an *i*th study), hat values (a leverage index that can be used to estimate the variance in the common effect size), as well as study weights (the inverse of precision when estimating the variance in the common effect size) after removing each study from the analysis were also examined in the conventional analysis (Viechtbauer & Cheung, 2010). All diagnostic estimates were obtained using the *influence* function in the R *metafor* package. Examining these plots allowed for a determination of the influential studies on estimating the common effect size and its variance.

In addition, a moderator analysis using no-intercept values was performed in the R *metafor* package for participant gender, client intervention site location, primary diagnostic category of participants, and research design. These moderator variables were selected because they were relevant to understanding the conditions under which VCT may be more or less effective and because inclusionary articles most consistently reported codable data; all other potential moderators had too much missing or unknown data to offer any meaningful implications. Each moderator's effect in the conventional analysis was first tested separately following the recommendation of Hox (2010), as cited in Assink and Wibbelink (2016), to avoid the problem of multicollinearity. Only statistically significant variables were considered in the final mixed-model.

Several common diagnostic statistical parameters were used to evaluate random-effects and mixed-effects (with the presence of moderators) models: QE (Cochran's *Q* - test) statistic or the amount of residual heterogeneity not accounted by the model (Shadish & Haddock, 2009): QM statistic for testing the null hypothesis that the moderator levels coefficients are equal in their respective values (e.g., $\beta_1 = \beta_2 = \beta_3$, etc.; Viechtbauer, 2010); I^2, a descriptive statistic that estimates the proportion of total variation (true effect size and sampling variation) that is due to heterogeneity among the effect sizes (Shadish & Haddock, 2009); random-effects variance or variance attributed to "true" differences in the effect sizes not due to sampling (τ^2) ($\tau^2 = 0$ implies $I^2 = 0$; Viechtbauer, 2010); and R^2 or a change in τ^2 due to the presence of moderators.

To specifically evaluate the variance components of the sampling as well as the outcome- and study-levels, a three-level Hierarchical Linear Model (HLM3) as implemented in R *metaSEM* (Cheung, 2015) package was used. This model was presented by Cheung (2014) and its applications in R libraries are described by Cheung (2019a, 2019b). The HLM3 approach to meta-analysis is particularly suitable in addressing stochastically-dependent outcomes. These dependencies can originate as a function of the same subjects being measured across multiple related outcomes (multiple-outcomes; e.g., BDI and PCL-5), the same subjects being measured on one outcome repeatedly (multiple-time points), or a combination of both.

The HLM3 model specifically addresses second-level (τ_2^2; outcomes nested within studies) and third-level (τ_3^2; studies) variance components in addition to addressing subjects' sampling variation (level-1). If the dependence between outcomes is ignored, the sampling variance (ε_{ij}) is not conditionally independent and the standard errors (*SE*s) are incorrect (typically too small), resulting in inferences that are consequently jeopardized. The HLM3 approach allows for the investigation of heterogeneity at the model specified levels of nesting. The HLM3 method, does not require a priori knowledge of the level of correlations/covariances among stochastically-dependent outcomes used to correct for the dependencies, such as in the case of manual adjustment in the conventional analysis or a robust error variance approach (Cheung, 2019a).

The studies and their corresponding individual outcomes were

[2] Hedges and Olkin (1985) on p. 211 illustrated that the weighted effect size can be computed from an estimated covariance matrix, $\widehat{\Sigma}^{(i)} = D_i R D_i$, where D_i is a diagonal matrix of *SD*s of the outcomes effect sizes and R is a correlation matrix of the effect sizes with the diagonal elements of 1 and off-diagonal elements as the estimated outcomes' correlations. The inverse $\widehat{\Lambda}^{(i)}$ of $\widehat{\Sigma}^{(i)}$ is computed. Next, a weighted estimator of the effect size can be obtained by $w_i = \frac{\widehat{\Lambda}^{(i)} e}{e' \widehat{\Lambda}^{(i)} e}$, where e is a *p*-dimensional column vector of 1 s. The weighted estimated of the effect size can be computed from $\widehat{\bar{\delta}}_i = w_i' d_i$, where d_i is a vector of outcome effect sizes. The final weighted estimate of the variance for the pooled effect size is $\widehat{\sigma}^2(\widehat{\bar{\delta}}) = \frac{1}{e' \widehat{\Lambda}^{(i)} e}$.

analyzed with a "meta3" function in the R *metaSEM* library with default setting, Maximum Likelihood estimation method (ML). A "study id" (level-3) served as a grouping element used to estimate the variance component between-studies (level-2). The first level captures the sampling variance within the outcomes, which is directly related to the number of participants associated with a particular outcome. Raudenbush (2009) and Raudenbush and Bryk (1985, 2002) refer to this component as a ν-known variance, an estimate that was originally proposed by Hedges (1981).

To specifically test the statistical significance of the source of heterogeneity (within- and between-studies), two models—one that did not include within-study variance or a model equivalent to conventional meta-analysis ($\tau^2{}_2 = 0$) and one that did not include between-study variance ($\tau^2{}_3 = 0$)—were built and compared to a full-model (in which both components of variance were freely estimated) with a procedure outlined in Cheung (2015).

The moderator analysis was also conducted using the HML3 approach with an intercept constraint equal to 0. Such a parameterization allows for the estimation of average effects of all levels of a moderator variable (Cheung, 2015).

The likelihood ratio (LR) test using a chi-square approximation for the change in model fit evaluated with a − 2 log likelihood (−2LL) function was used to determine both constrained variance models versus a freely estimated model (relevant to testing the statistical significance of the sources of heterogeneity) as well in testing models that included moderators versus a base model (without moderators).

3.2. Reliability studies

Fourteen studies (see Table 8) examining the reliability of assessment outcomes across VCT and in-person examiners were included for analysis. The number of reported outcomes (i.e., diagnostic decisions, conclusions, or results following from the assessment or evaluation) varied from one to 16 and the weighted average number of clients assessed per outcome was 27.8. All outcomes used dependent-samples when evaluating clients in VCT and in-person assessment conditions. Of relevance in the reliability assessments was not the intervention effect, but rather whether or not the assessment format (VCT or in-person) was associated with differences in the reported outcomes of the respective assessment. Most studies reported means and *SD*'s or *t*-values for dependent samples with a corresponding number of participants and/or *p*-values. Grob, Weintraub, Sayles, Raskin, and Ruskin (2001) reported only within-groups reliability coefficients. All available outcome statistics were converted into Hedges' *g* standardized effect sizes and their corresponding variances using effect sizes formulas for dependent-samples presented by Borenstein (2009) with a correlation coefficient of 0.5.

4. Results

4.1. Intervention studies

4.1.1. Study and sample characteristics

Forty-three individual studies published between the years of 2002 to 2019 met all inclusionary criteria and were included in the final analyses examining intervention outcomes. Seven studies were published in 2016; five in 2017; four in 2015 and between one and three studies were published in other years. A total of 281 individual outcomes were included across these 43 studies. Methodological features and sample characteristics of the analyzed intervention studies are presented in Tables 1 and 2, respectively.

The final analysis of interventions outcomes was based on a total of 4336 participants with a simple mean age of $M = 38.0$ years ($SD = 16.1$) and a sample size weighted age of $M = 47.0$ years. The average number of analyzed outcomes per study was six and a half and the average numbers of participants in the VCT and in-person conditions were 47.2 and 53.6, respectively. Twenty studies (consisting of 36.3% of the outcomes analyzed) included both female and male clients, 13 studies (26.7% of outcomes) included mostly males,[3] five studies (16.7% of outcomes) included males only, four studies (19.2% of outcomes) included mostly females, and finally one study (1.1% of outcomes) included females only. Regarding the location of clients, 12 studies (26.3% of outcomes) reported on services provided in their homes or a university clinic, nine studies (18.1% of outcomes) in Veteran hospitals/facilities, seven studies (15.7% of outcomes) in medical facilities, and also seven studies (11.0.0% of outcomes) in outpatient clinics, two studies (6.1% of outcomes) in correctional facilities, and six (28.8% of outcomes) in other or not reported types of facilities. Trauma and anxiety related symptoms were measured in 15 studies (34.9% of outcomes), depressive or mood disorders were measured in 15 studies (25.3% of outcomes), and other conditions were examined in 13 studies (one to four studies per additional diagnostic category 39.8% of outcomes), including three that did not report any specific diagnostic criteria or treatment target for participant inclusion. Twenty-seven studies (71.5% of the total number of outcomes analyzed) employed primarily psychosocial interventions or psychotherapy, 13 studies (26.3% of outcomes) focused primarily on psychotropic medication management with supplemental counseling or psychoeducational services, and three studies (2.1% of outcomes) indicated implementing a psychological or psychiatric intervention but did not report further details on the services provided. Study specific counseling approaches are detailed in Table 2 for studies that reported this information.

Thirty-three studies (82.9% of the outcomes analyzed) used a random assignment procedure, eight studies used alternative assignment designs, such as a matched design (15.3% of outcomes), and two did not report the method of condition assignment (1.8% of outcomes). Five of the 43 studies (11.6%) involved child or adolescent clients (between ages 3 and 17 years old across studies). Some (e.g., Himle et al., 2012), treated the child and measured outcomes based on both provider and caregiver report; others (e.g., Tse, McCarty, Stoep, & Myers, 2015) treated both the child and caregiver.

4.1.2. Conventional meta-analysis

Aggregate Between-Group Comparison. Consistent with our general hypothesis, the overall effect size for the estimated model was not statistically significant, Hedges' $g = -0.02$, 95% CIs [−0.12, 0.94], ($SE = 0.06$), $p = .788$, indicating no statistically significant aggregate difference between outcomes associated with videoconference-delivered interventions and those associated with in-person interventions. The estimated total heterogeneity or τ^2 was 0.09 (SE = 0.03). A statistically significant amount of heterogeneity was observed in the model, $Q(42) = 160.41$, $p < .001$. The ratio of studies' heterogeneity to total observed variance—a descriptive statistic, expressed as I^2 was 72.52%.

Fig. 2 presents the forest plot under a random-effects model. The plot includes each study's effect size (Hedges' *g*), standard error (*SE*), variance, 95% confidence interval for the effect size, as well as associated *z*- and *p*-values for the hypothesis that the effect size of a given study is equal to zero. In Fig. 2, it can be observed that the overall effect size is estimated with a high precision—that is, there are small confidence intervals around the mean effect size values. Studies that included a higher number of participants generally demonstrated smaller bands of confidence intervals than studies that included relatively few participants (see Table 3 for participant numbers). Gros et al. (2011) stands out as being associated with an unusually high effect size favoring the in-person treatment condition.

Analysis of Outliers. Results of standardized residuals, DFFITS, Cook's values, covariance ratios, variance (τ^2), hat values, and weights when each study is removed from the analysis in turn are presented in Fig. 4. The plots in the figure identify study #17 (Gros et al., 2011), #22

[3] "Mostly" is defined as making up more than 80% of the total sample within a given study.

Table 2
Sample descriptive statistics of included studies (Intervention Studies).

Study	Number of outcomes	Ave. *n* VCT	Ave. *n* In-Person	Client Gender Composition	Client Racial Composition	Mean Age of Clients	Primary Client Diagnostic Category	Intervention Approach
Acierno et al. (2016)	7	45	46	Mostly male	Multicultural	46	Trauma and Anxiety Related	Individual Behavioral Activation
Batastini and Morgan (2016)	13	24	12	Male only	Multicultural	30	NA	Group Cognitive-Behavioral
Bouchard et al. (2004)	16	11	10	Mixed	NA	NA	Trauma and Anxiety Related	Individual Cognitive-Behavioral
Choi et al. (2014)	6	43	50	Mixed	Multicultural	65	Depressive or Mood Disorder	Individual Problem-Solving
Chong & Moreno et al. (2012)	6	64	68	Mostly female	NA	NA	Depressive or Mood Disorder	NA
Comer, Furr, Kerns, et al. (2017)	10	11	10	Mixed	Mostly Caucasian	7	Obsessive-Compulsive and Related	Individual Cognitive-Behavioral
Comer, Furr, Miguel, et al., 2017)	13	17	15	Mostly male	Multicultural	4	Disruptive, Impulse-Control, and Conduct Disorders	Parent-Child Interaction
Day and Schneider (2002)	4	26	27	Mixed	Mostly Caucasian	39	NA	NA
De la Cuevas et al., 2006	4	66	64	Mixed	NA	NA	Trauma and Anxiety Related	Individual Cognitive-Behavioral
Farabee et al. (2016)	1	20	40	Mixed	Multicultural	38	Other	NA
Fortney et al. (2007)	8	177	218	Mostly male	Multicultural	59	Depressive or Mood Disorder	NA
Fortney et al. (2011)	6	153	195	Mostly male	Multicultural	59	Depressive or Mood Disorder	Individual Cognitive-Behavioral
Fortney et al. (2013)	12	141	155	Mostly female	Multicultural	47	Depressive or Mood Disorder	Individual Cognitive-Behavioral
Frueh et al. (2007)	12	8	11	Male only	Multicultural	NA	Trauma and Anxiety Related	Group Cognitive-Behavioral
Germain et al. (2009)	4	16	32	Mixed	NA	NA	Trauma and Anxiety Related	Individual Cognitive-Behavioral
Glassman et al. (2019)	6	45	47	Mixed	NA	NA	Trauma and Anxiety Related	Individual Cognitive Processing
Gros et al. (2011)	2	30	27	Mostly male	Multicultural	NA	Trauma and Anxiety Related	Individual Cognitive-Behavioral/Exposure
Herbert et al. (2017)	6	63	65	Mostly male	Multicultural	52	Other Conditions of Clinical Attention	Individual Acceptance-Commitment
Himle et al. (2012)	6	10	8	Mostly male	NA	12	Neurocognitive Disorders	Individual Behavioral
Hulsbosch et al. (2017)	12	34	38	Mixed	NA	46	Schizophrenia Spectrum and Other Psychotic	Individual Acceptance-Commitment
Hungerbuehler et al. (2016)	8	47	42	Mixed	Multicultural	36	Depressive or Mood Disorder	NA
Kelleher et al. (2019)	2	66	67	Mixed	Multicultural	56	Other Conditions of Clinical Attention	Individual Cognitive-Behavioral
King et al. (2014)	1	24	35	Mixed	Multicultural	41	Other Conditions of Clinical Attention	NA
Liu et al. (2019)	6	73	68	Mixed	Multicultural	48	Trauma and Anxiety Related	Individual Cognitive Processing
Luxton, Pruitt, et al. (2016)	6	39	40	Mostly male	Multicultural	NA	Depressive or Mood Disorder	Individual Behavioral Activation
Maieritsch et al. (2016)	12	25	26	Mostly male	NA	31	Trauma and Anxiety Related	Individual Cognitive Processing
Mitchell et al. (2008)	33	35	33	Mostly female	Mostly Caucasian	30	Feeding and Eating	Individual Cognitive-Behavioral
Modai et al. (2006)	1	39	42	Mixed	NA	NA	Schizophrenia Spectrum and Other Psychotic	NA
Moreno et al. (2012)	3	68	72	Mostly female	NA	NA	Depressive or Mood Disorder	NA
Morgan et al. (2008)	4	43	50	Male only	Multicultural	32	Depressive or Mood Disorder	NA
Morland et al. (2010)	10	61	64	Male only	Multicultural	NA	Trauma and Anxiety Related	Group Anger Management
Morland et al. (2014)	8	56	59	Male only	Multicultural	46	Trauma and Anxiety Related	Individual Cognitive Processing
Morland et al. (2015)	3	41	45	Female only	Multicultural	46	Trauma and Anxiety Related	Individual Cognitive-Processing
Nelson et al. (2003)	1	14	14	Mixed	Multicultural	10	Depressive or Mood Disorder	Individual Cognitive-Behavioral
O'Reilly et al. (2007)	2	181	197	Mixed	NA	NA	Depressive or Mood Disorder	NA
Ruskin et al. (2004)	1	59	60	Mostly male	Multicultural	50		NA

(*continued on next page*)

Table 2 (*continued*)

Study	Number of outcomes	Ave. *n* VCT	Ave. *n* In-Person	Client Gender Composition	Client Racial Composition	Mean Age of Clients	Primary Client Diagnostic Category	Intervention Approach
							Depressive or Mood Disorder	
Shulman et al. (2017)	1	11	11	Mixed	Multicultural	40	Depressive or Mood Disorder	NA
Simpson et al. (2015)	1	6	17	Mixed	NA	NA	Depressive or Mood Disorder	Individual Cognitive-Behavioral
Stubbings et al. (2013)	6	10	9	Mixed	Multicultural	30	Depressive or Mood Disorder	Individual Cognitive-Behavioral
Tse et al. (2015)	10	12	25	Mixed	Mostly Caucasian	9	Other Conditions of Clinical Attention	Individual Behavioral
Tuerk et al. (2010)	2	9	29	Mostly male	Multicultural	39	Trauma and Anxiety Related	Individual Prolonged Exposure
Wierwille et al. (2016)	2	85	136	Mostly male	Mostly Caucasian	47	Trauma and Anxiety Related	Individual Cognitive Processing
Yuen et al. (2015)	4	23	29	Mostly male	Multicultural	44	Trauma and Anxiety Related	Individual Prolonged Exposure

Note: "Mixed" gender and "multicultural" race/ethnicity indicated that no one category within these variables was reported as over 80% of the sample. "Other Conditions That May Be a Focus of Clinical Attention" was a coding category that was designated for that did not occur with enough frequency to parse into a separate diagnostic category or for which detailed information was not reported. "Mostly female" gender also included one study (Morland et al., 2015) that treated all females. These categories were combined to allow for inclusion in the moderator analyses and reduce statistical limitations due to small group sizes.

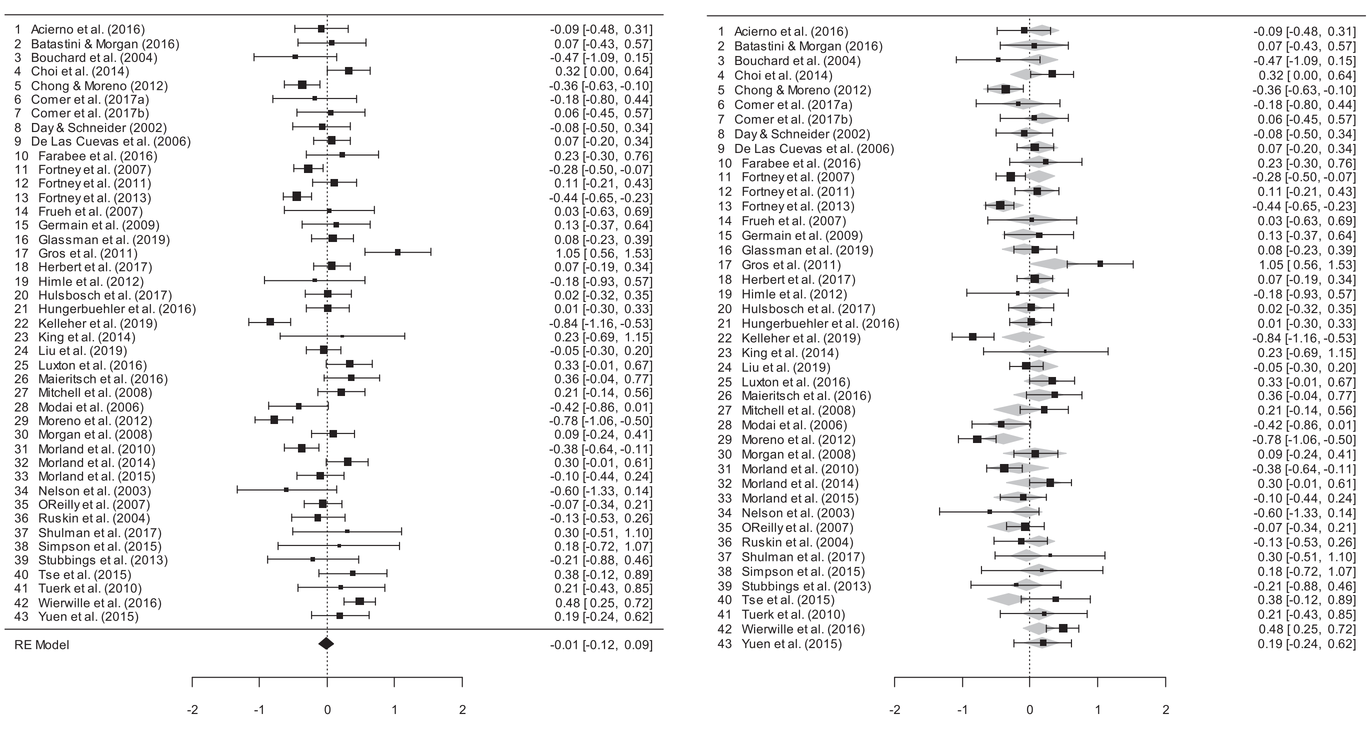

Fig. 2. Forest plot to the left demonstrates the results for 43 studies that examined the effects of VCT and in−person treatments using non − nested design with variance adjustments for studies as presented by Hedges and Olkin (1985). The figure shows estimated Hedges' *g* effect sizes and their corresponding *SE*s and 95% confidence intervals (the far−right column). The forest plot to the right demonstrates the same effect sizes with grayed areas showing the results corrected by the presence of a statistically significant unordered factor moderator: *participant gender* and *client intervention site*, which had a statistically significant effect for medical facility location subgroup.

(Kelleher et al., 2019) and #29 (Moreno, Chong, Dumbauld, Humke, & Byreddy, 2012) as outliers, as evidenced by the following diagnostic parameters for both studies, *r* studentized residuals exceeding the value of 1.96, covariance ratio falling below the value of 1 and more pronounced leverage Cook's or leverage values. These studies, however, do not substantially change the overall model fit as demonstrated by the two bottom plots in Fig. 4 (hat and weight values). In addition, the results of a leave-one-out study analysis (see Table 6) demonstrated and confirmed that leaving out any one of the 43 studies did not meaningfully change the overall Hedges' *g* estimates or the corresponding *z* and *p* values. Therefore, all studies were retained.

Moderator Analysis. The following moderators (all represented as unordered factors) were considered: *primary client gender* (levels*: males*, *mostly males*, *mostly females*, and *mixed*); *research procedures* (levels:

Table 3
Intervention studies and corresponding outcomes with average number of participants in treatment conditions (VCT and In-Person).

Study	Number of outcomes per study	Means and dispersion	Participant counts
Acierno et al. (2016)	7	7	–
Batastini and Morgan (2016)	13	13	–
Bouchard et al. (2004)	16	16	–
Choi et al. (2014)	6	6	–
Chong & Moreno et al. (2012)	6	6	–
Comer, Furr, Kerns, et al. (2017)	10	10	–
Comer, Furr, Miguel, et al. (2017)	13	13	–
Day and Schneider (2002)	4	4	–
De la Cuevas et al., 2006	4	4	–
Farabee et al. (2016)	1	1	–
Fortney et al. (2007)	6	2	4
Fortney et al. (2011)	8	4	4
Fortney et al. (2013)	12	6	6
Frueh et al. (2007)	12	12	–
Germain et al. (2009)	4	4	–
Glassman et al. (2019)	6	6	–
Gros et al. (2011)	2	2	–
Herbert et al. (2017)	6	6	–
Himle et al. (2012)	6	4	2
Hulsbosch et al. (2017)	12	12	–
Hungerbuehler et al. (2016)	8	8	–
Kelleher et al. (2019)	2	2	–
King et al. (2014)	1	–	1
Liu et al. (2019)	6	6	–
Luxton, Nelson, and Maheu (2016)	6	6	–
Maieritsch et al. (2016)	12	12	–
Mitchell et al. (2008)	33	27	6
Modai et al. (2006)	1	1	–
Moreno et al. (2012)	3	3	–
Morgan et al. (2008)	4	4	–
Morland et al. (2010)	10	10	–
Morland et al. (2014)	8	8	–
Morland et al. (2015)	3	3	–
Nelson et al. (2003)	1	1	–
O'Reilly et al. (2007)	2	1	1
Ruskin et al. (2004)	1	–	1
Shulman et al. (2017)	1	1	–
Simpson et al. (2015)	1	1	–
Stubbings et al. (2013)	6	6	–
Tse et al. (2015)	10	10	–
Tuerk et al. (2010)	2	2	–
Wierwille et al. (2016)	2	2	–
Yuen et al. (2015)	4	4	–
All studies	281	256	25

Note. Studies are identified by first author and year. The average number of clients per condition is reported–some studies had attrition of subjects in repeated measurement of outcomes.

random assignment, *non-random assignment* and *not reported*); *client intervention site* (e.g., where clients went to receive services; levels: *correctional facility/prison/jail* (including internal psychiatric unit), *in-home or university clinic*, *medical facilities*, *outpatient psych clinic*, *Veteran hospital/facility*, and *not reported*); as well as the *primary clinical diagnostic category* of focus (levels: *depressive or mood disorder*; *disruptive, impulse-control, and conduct disorders*; *feeding and eating disorders*; *neuro-cognitive disorders*, *obsessive-compulsive and related disorders*; *schizophrenia spectrum and other psychotic disorders*; *trauma and anxiety related* (to include PTSD); *other conditions of clinical attention*; and *not reported*).

Of note, due to the inconsistent reporting of data within articles and the need to obtain a sufficient count within each moderator level, originally coding categories for *client intervention site* and *primary diagnostic category* were collapsed to form the above categories. For example, *medical facility* included community medical facilities and general medical centers; *outpatient psych clinic* included substance dependency and psychiatric treatment facilities; *in-home or university clinic* sites were combined based on a general assumption that clients being seen in these settings likely share similar risk/severity levels and were not part of another specialized population (e.g., veterans). In many studies, the primary diagnostic focus of treatment was mixed. In these cases, the highest reported frequency was used to categorize this variable (e.g., one study included multiple DSM disorders but schizophrenia disorders were prevalent among 86% of the VCT group and 64% of the in-person group, therefore, this study was classified under *schizophrenia spectrum and other psychotic disorders*; Modai et al., 2006) Although efforts were made to maintain some degree of specificity whenever possible, several outcomes were too few and too distinct to theoretically justify collapsing under a more specific label; these outcomes made up an *other conditions of clinical attention* category. This category included substance and addictive disorders, ADHD, and chronic pain (addressed via a behavioral health intervention). Studies in which the outcome of interest was not explicitly described in the study or could not be assumed were captured by a *not reported* category. Table 2 shows these moderator categories by included study.

Test statistics for residual heterogeneity (QE) and corresponding degrees of freedom (*df*) and *p*-values, test statistics for moderators effects (QE) and corresponding degrees of freedom (*df*) and *p*-values, residual heterogeneity (τ^2) with *SE*, the ratio of studies' heterogeneity to total observed variance I^2, and the amount of residual variance explained by moderators (R^2) associated with each moderator variable are reported in Table 4.

Primary client gender and *client intervention site* were statically significant predictors of treatment effectiveness by modality, $Q = 8.97$, $p = .030$ and $Q = 11.17$, $p = .048$, respectively. The amount of heterogeneity accounted for was $R^2 = 19.80$ for gender and 23.72 for client intervention site. The effects of the levels of moderators (respective levels mean effect sizes, *SE*s, and corresponding *z* and *p* values) as unordered factors in a model that did not include an intercept value—to estimate the mean value (not a level's displacement relative to the intercept value) of each level of the moderator are summarized in Table 5. Regarding *gender*, the coefficient for *mostly females,* $\beta_1 = -0.32$ (favoring VCT) was statistically significant, $p = .020$. The coefficient for mostly males was significant at $\alpha = 0.10$, $\beta_3 = 0.16$, $p = .076$ (favoring in-person). With respect to the *client intervention site*, clients seen in medical facilities reported a statistically significantly more favorable response to VCT interventions than in-person, $\beta_3 = -0.33$, $p = .004$. No other moderator variables were statistically significant.

Analysis of Publication Bias. Although publication bias is arguably less of a concern in this line of research, as non-significant findings are commonly published in support of the feasibility and effectiveness of telehealth services compared to in-person services, results of the analysis of publication bias for random- and mixed-effects models are presented in Fig. 3. These plots depict the *SE*s the Y-axis as a function of observed effect sizes on the X-axis. In the left grayed area of the top left plot, 6 studies have relatively small *SE*s and unusually high effect sizes favoring VCT interventions. On the right, the plot depicts two studies (one with a small and one with a larger *SE*) that favor in-person interventions. The other plots in Fig. 3 depict the publication bias for mixed-effects models, where the presence of the moderators demonstrates that all points are contained within the triangle of expected outcomes, showing no evidence for publication bias.

4.1.3. Three-level hierarchical linear model (HLM3)

Aggregate Between-Group comparison. The overall effect size estimated by the HLM3 model was not statistically significant (i.e., no difference between VCT and in-person treatments was observed), Hedges' $g = -0.02$, 95% CIs [−0.11, 0.07], ($SE = 0.05$), $p = .682$. A statistically significant amount of heterogeneity was observed, $Q(280) = 1131.37$, $p < .001$. These results corresponded very closely to results of the conventional analysis.

The levels 2 and 3, I^2 and τ^2 were estimated using a likelihood-based interval, which is a more accurate method of obtaining the CIs than the conventional Wald statistics (Cheung, 2019b). The τ^2_2 and τ^2_3were 0.11

Table 6
Intervention studies Leave-One-Out Effect Sizes (Hedges' g) and corresponding statistics as scaled by a random-effects model.

Study	Hedges' *g*	*z*	*p*	95% CI LL UL	Q	p_Q	τ^2	I^2
1. Acierno et al. (2016)	−0.01	−0.23	0.821	[−0.13, 0.10]	160.38	> 0.001	0.09	73.23
2. Batastini and Morgan (2016)	−0.02	−0.29	0.772	[−0.13, 0.10]	160.17	> 0.001	0.09	73.27
3. Bouchard et al. (2004)	−0.01	−0.13	0.897	[−0.12, 0.10]	158.68	> 0.001	0.09	72.83
4. Choi et al. (2014)	−0.03	−0.44	0.664	[−0.14, 0.09]	154.87	> 0.001	0.09	72.31
5. Chong & Moreno et al. (2012)	0.00	−0.08	0.939	[−0.12, 0.11]	154.78	> 0.001	0.09	71.99
6. Comer, Furr, Kerns, et al. (2017)	−0.01	−0.21	0.832	[−0.12, 0.10]	160.25	> 0.001	0.09	73.23
7. Comer, Furr, Miguel, et al. (2017)	−0.02	−0.29	0.774	[−0.13, 0.10]	160.21	> 0.001	0.09	73.27
8. Day and Schneider (2002)	−0.01	−0.23	0.818	[−0.13, 0.10]	160.39	> 0.001	0.09	73.26
9. De la Cuevas et al., 2006	−0.02	−0.30	0.762	[−0.13, 0.10]	159.50	> 0.001	0.09	72.88
10. Farabee et al. (2016)	−0.02	−0.35	0.727	[−0.13, 0.09]	159.28	> 0.001	0.09	73.08
11. Fortney et al. (2007)	−0.01	−0.11	0.913	[−0.12, 0.11]	155.86	> 0.001	0.09	71.96
12. Fortney et al. (2011)	−0.02	−0.32	0.751	[−0.13, 0.09]	159.39	> 0.001	0.09	73.02
13. Fortney et al. (2013)	0.00	−0.02	0.982	[−0.11, 0.11]	146.09	> 0.001	0.08	70.81
14. Frueh et al. (2007)	−0.02	−0.28	0.784	[−0.13, 0.10]	160.34	> 0.001	0.09	73.26
15. Germain et al. (2009)	−0.02	−0.32	0.753	[−0.13, 0.09]	159.86	> 0.001	0.09	73.21
16. Glassman et al. (2019)	−0.02	−0.30	0.762	[−0.13, 0.10]	159.68	> 0.001	0.09	73.04
17. Gros et al. (2011)	−0.04	−0.74	0.459	[−0.14, 0.06]	140.57	> 0.001	0.07	68.51
18. Herbert et al. (2017)	−0.02	−0.30	0.763	[−0.13, 0.10]	159.47	> 0.001	0.09	72.84
19. Himle et al. (2012)	−0.01	−0.22	0.824	[−0.12, 0.10]	160.30	> 0.001	0.09	73.22
20. Hulsbosch et al. (2017)	−0.02	−0.27	0.786	[−0.13, 0.10]	160.23	> 0.001	0.09	73.16
21. Hungerbuehler et al. (2016)	−0.02	−0.27	0.787	[−0.13, 0.10]	160.22	> 0.001	0.09	73.10
22. Kelleher et al. (2019)	0.01	0.13	0.894	[−0.10, 0.11]	135.10	> 0.001	0.07	67.79
23. King et al. (2014)	−0.02	−0.31	0.757	[−0.13, 0.09]	160.04	> 0.001	0.09	73.13
24. Liu et al. (2019)	−0.01	−0.24	0.814	[−0.13, 0.10]	160.41	> 0.001	0.09	72.85
25. Luxton et al. (2016)	−0.02	−0.43	0.665	[−0.14, 0.09]	155.26	> 0.001	0.09	72.37
26. Maieritsch et al. (2016)	−0.02	−0.43	0.667	[−0.13, 0.09]	156.30	> 0.001	0.09	72.50
27. Mitchell et al. (2008)	−0.02	−0.37	0.714	[−0.13, 0.09]	158.18	> 0.001	0.09	72.87
28. Modai et al. (2006)	−0.01	−0.10	0.922	[−0.12, 0.11]	157.64	> 0.001	0.09	72.58
29. Moreno et al. (2012)	0.01	0.12	0.902	[−0.10, 0.11]	133.54	> 0.001	0.07	67.98
30. Morgan et al. (2008)	−0.02	−0.31	0.759	[−0.13, 0.09]	159.65	> 0.001	0.09	73.06
31. Morland et al. (2010)	0.00	−0.07	0.943	[−0.11, 0.11]	154.38	> 0.001	0.09	71.93
32. Morland et al. (2014)	−0.02	−0.42	0.672	[−0.14, 0.09]	155.19	> 0.001	0.09	72.37
33. Morland et al. (2015)	−0.01	−0.22	0.828	[−0.12, 0.10]	160.34	> 0.001	0.09	73.15
34. Nelson et al. (2003)	−0.01	−0.13	0.900	[−0.12, 0.10]	158.32	> 0.001	0.09	72.75
35. OReilly et al. (2007)	−0.01	−0.23	0.819	[−0.13, 0.10]	160.40	> 0.001	0.09	72.97
36. Ruskin et al. (2004)	−0.01	−0.21	0.836	[−0.12, 0.10]	160.26	> 0.001	0.09	73.20
37. Shulman et al. (2017)	−0.02	−0.33	0.740	[−0.13, 0.09]	159.68	> 0.001	0.09	73.09
38. Simpson et al. (2015)	−0.02	−0.30	0.764	[−0.13, 0.09]	160.15	> 0.001	0.09	73.16
39. Stubbings et al. (2013)	−0.01	−0.21	0.835	[−0.12, 0.10]	160.20	> 0.001	0.09	73.21
40. Tse et al. (2015)	−0.02	−0.41	0.680	[−0.13, 0.09]	157.49	> 0.001	0.09	72.69
41. Tuerk et al. (2010)	−0.02	−0.33	0.744	[−0.13, 0.09]	159.76	> 0.001	0.09	73.15
42. Wierwille et al. (2016)	−0.03	−0.57	0.567	[−0.14, 0.08]	139.49	> 0.001	0.08	70.38
43. Yuen et al. (2015)	−0.02	−0.35	0.729	[−0.13, 0.09]	159.10	> 0.001	0.09	73.06

Note. Positive Hedges' *g* estimate implies an advantage of in-person interventions over VCT.

Table 4
Results of moderator analysis and corresponding statistics modeled by conventional mixed-effects model: Overall model statistics.

	QE	df_{QE}	p_{QE}	QM	df_{QM}	p_{QM}	τ^2	$SE_{\tau 2}$	I^2	R^{2*}
Gender	120.71	39	< 0.001	8.97	3	0.030	0.07	0.03	66.99	19.80
Client intervention site	107.87	37	< 0.001	11.17	5	0.048	0.07	0.03	65.86	23.72
DSM disorder	133.23	34	< 0.001	4.94	8	0.764	0.09	0.03	74.31	0.00
Research design	142.21	40	< 0.001	2.11	2	0.35	0.09	0.03	71.76	2.17
Gender + Client intervention site	101.79	34	< 0.001	13.94	8	0.08	0.07	0.03	65.81	21.10

with 95% CIs [0.08, 0.15] and 0.05 with 95% CIs [0.02, 0.10], respectively. The I^2_2 and I^2_3, using the same method, were 0.53 with 95% CIs [0.53, 0.65] and 0.23 with 95% CIs [0.10, 0.40], respectively. Therefore, the total variation can be partitioned into 53.18% due to level-2 variance (outcomes) and 23.40% due to level-3 variance (studies). The remaining 23.42% of variance is due to level-1 variance or sampling of the participants within outcomes.

To evaluate the significance level of the two sources of heterogeneity (within- and between-studies), two models—one that did not include within-study variance or a model equivalent to conventional meta-analysis ($\tau^2_2 = 0$) and one that did not include between-study variance ($\tau^2_3 = 0$)— were evaluated and compared against a full-model. In both cases the full model ($-2LL = 385.10$) outperformed the no-between study variance model ($-2LL = 410.55$), $\chi^2(1) = 25.45$, $p < .001$, and the no-within study variance model ($-2LL = 537.17$), $\chi^2(1) = 188.07$, $p < .001$. The significant results of the constrained models versus the non-constrained variance model implied that the estimation of both variance components was justified.

Moderator Analyses. The same moderator variables used in the conventional analyses were likewise entered in the HLM3 models.

Table 5
Results of moderator analysis and corresponding statistics as scaled by conventional mixed-effects model: Variable levels.

Moderator	*k*	Hedges' *g*	*SE*	*z*	*p*	95% CI LB	UB
Gender							
Females mostly*, B_1	57	−0.32	0.14	−2.33	0.020	−0.58	−0.05
Males, B_2	47	0.01	0.15	0.07	0.945	−0.28	0.30
Males mostly, B_3	75	0.16	0.09	1.78	0.076	−0.02	0.35
Mixed, B_4	102	−0.06	0.08	−0.70	0.484	−0.21	0.10
Client intervention site							
Correctional facility/prison/jail*, B_1	17	0.08	0.23	0.34	0.735	−0.38	0.54
Home or university clinic, B_2	74	0.04	0.10	0.43	0.666	−0.16	0.24
Medical facility, B_3	44	−0.33	0.11	−2.87	0.004	−0.55	−0.10
Outpatient clinic, B_4	31	0.17	0.14	1.21	0.227	−0.10	0.44
Veteran hospital, B_5	51	0.10	0.11	0.92	0.357	−0.11	0.30
Not reported*, B_6	64	−0.09	0.14	−0.66	0.509	−0.36	0.18
DSM disorder							
Depressive or mood disorder, B_1	71	−0.12	0.09	−1.28	0.202	−0.31	0.06
Disruptive, impulse-control, and conduct disorders, B_2	13	0.06	0.40	0.15	0.878	−0.72	0.84
Feeding and eating disorders, B_3	33	0.21	0.35	0.60	0.550	−0.48	0.90
Neurocognitive disorders, B_4	6	−0.18	0.49	−0.37	0.710	−1.14	0.77
Obsessive-compulsive and related disorders, B_5	10	−0.18	0.44	−0.41	0.682	−1.04	0.68
Other conditions of clinical attention, B_6	19	−0.12	0.19	−0.65	0.516	−0.50	0.25
Schizophrenia spectrum and other psychotic disorders, B_7	13	−0.19	0.26	−0.73	0.466	−0.69	0.31
Trauma and anxiety related, B_8	98	0.12	0.09	1.29	0.199	−0.06	0.31
Not reported*, B_9	18	0.06	0.23	0.28	0.777	−0.38	0.51
Research design							
Non-random, B_1	43	0.10	0.14	0.74	0.460	−0.17	0.38
Random, B_2	233	−0.06	0.06	−0.88	0.376	−0.18	0.07
Not reported*, B_9	5	0.21	0.23	0.93	0.354	−0.24	0.67

Note. Positive estimates imply an advantage of in-person interventions over VCT. k = number of samples, CI = confidence interval, LL = lower limit, UL = upper limit.

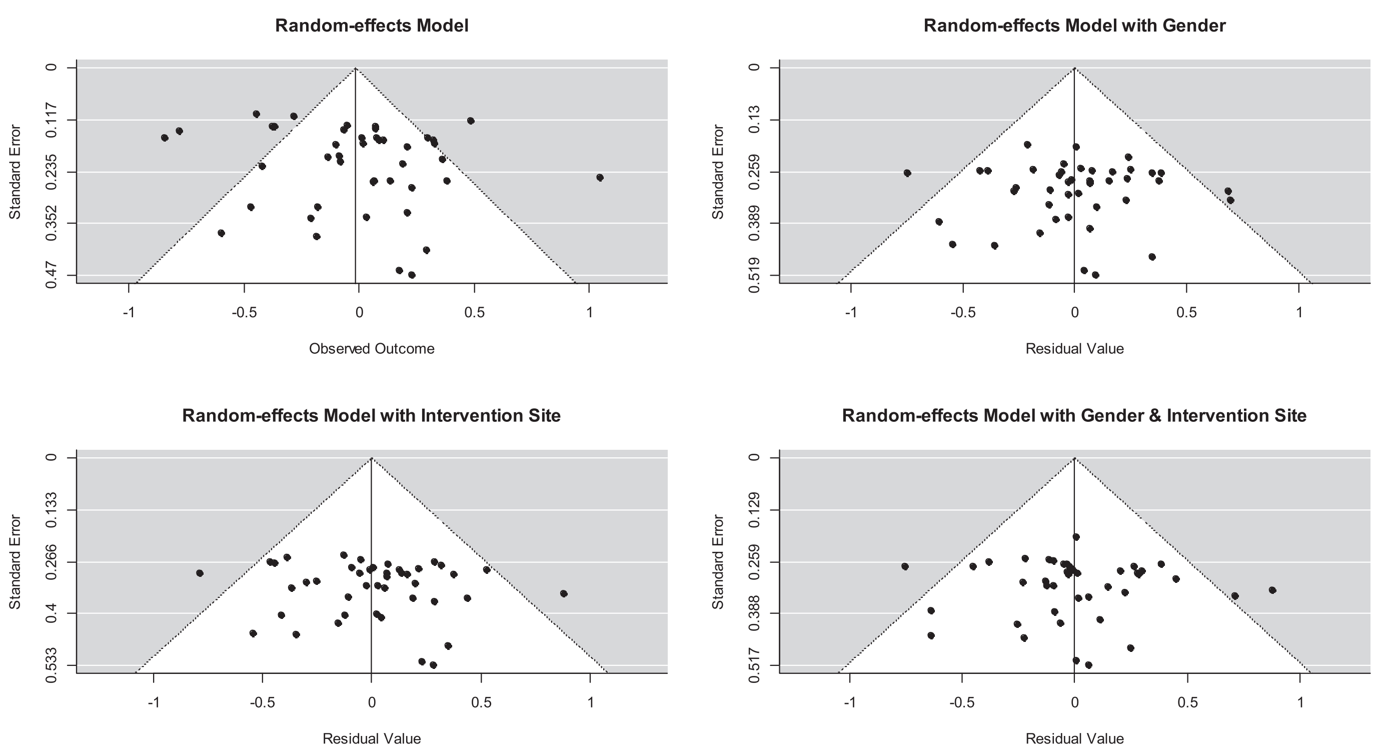

Fig. 3. Funnel plot of standardized residuals *SEs* for a random−effect model (upper left panel), mixed−effect model with a *gender* moderator (upper right panel), mixed−effect model with a *client intervention site* moderator (lower left panel), and a mixed−effect model with both moderators (lower right panel). A correction to the publication bias (outlier studies) can be observed.

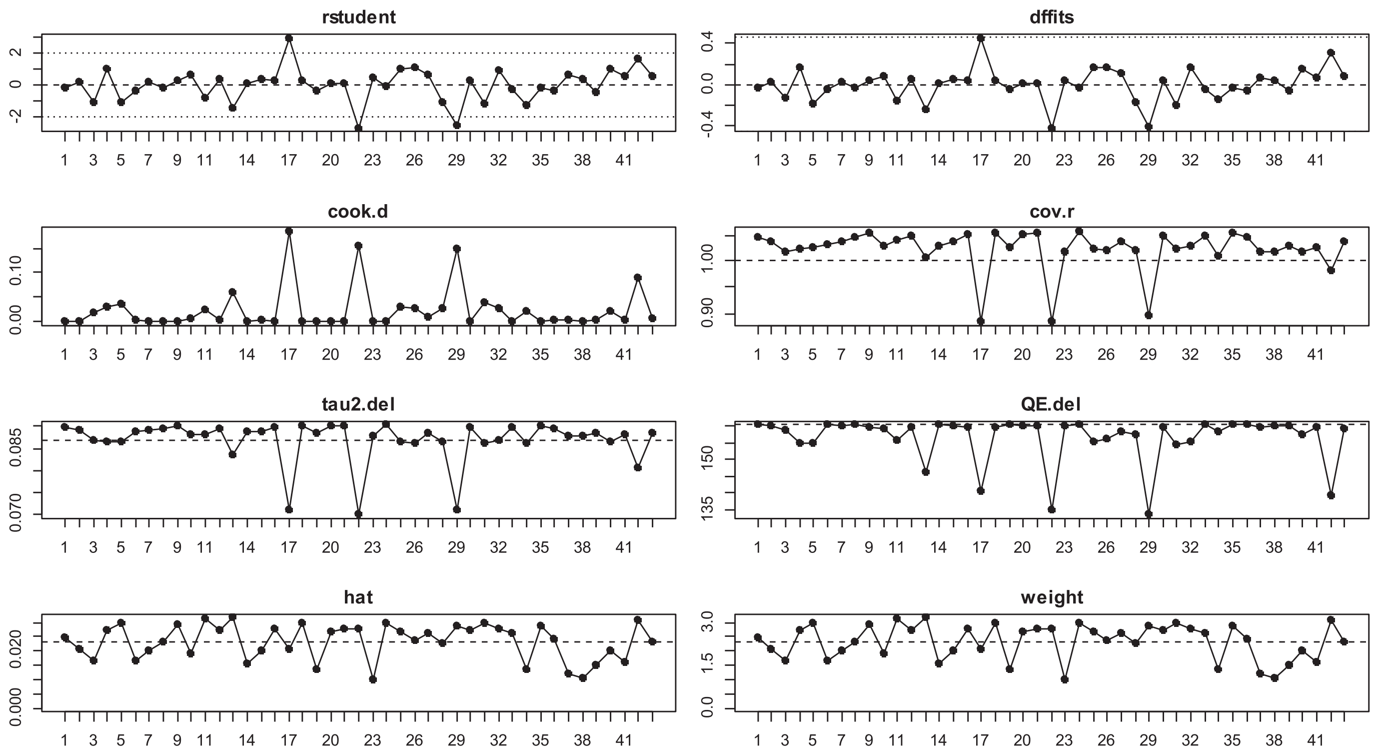

Fig. 4. Plots of standardized residuals, DFFITS and Cook's values, covariance ratios, variance (τ^2) as well as hat values and weights when each study is removed from the model. Study #17 (Gros et al., 2011), #22 (Kelleher et al., 2019) and #29 (Moreno et al., 2012) demonstrate high residual structures, but do not have a high influence on the overall model based on the hat and weight values.

Table 7
Results of moderator analysis and corresponding statistics as modeled by HLM3 mixed-effects model.

t1

Note. Positive Hedges' *g* estimate implies an advantage of in-person interventions over VCT.

Moderator models were assessed specifically to exact the effects of heterogeneity reduction at the outcome- and study-levels. The results of the HLM3 moderator analysis are presented in Table 7. One regression slope within the *participant gender* variable and one regression slope within the *client intervention site* variable were statistically significant. Specifically, the slope associated with the *mostly females* gender category was statistically significant, $\beta_1 = -0.24$ (favoring VCT), $p = .031$. The variance components for both level-2 and level-3 are included when the *gender* moderator also presented in the table demonstrate a reduction in outcome- and study-levels components. A total of 17.4% reduction in model heterogeneity can be explained at the study-level (i.e., between-studies) due to *gender* and *client intervention site*. The reduction in modeled variables at the outcome-level (i.e., between-outcomes) was negligible (2.4%). Therefore, gender effects were found to be rather dissimilar across different studies and are best predicted by the specific study in which they were measured (as opposed to how they were measured or what was being measured).

With respect to the *client intervention site*, clients treated in medical facilities again showed statistically significantly more favorable outcomes following VCT interventions than in-person intervention, $\beta_3 = -0.28$, $p = .004$. A total of 32.3% reduction in model heterogeneity can be explained specifically at the study-level due to *client intervention site*. The reduction in modeled variables at the outcome-level was negligible at 0.9%. Thus, it appears that site type effects are more so study-dependent than outcome-dependent.

When both significant moderators, *participant gender* and *client intervention site*, were considered simultaneously, a 32.7% reduction in model heterogeneity between-studies was observed. The reduction in variance between-outcomes nested in specific studies remained negligible at 2.4%. This between-outcomes finding is likely due to the fact that, typically, the same number of participants within a study were assessed at multiple-outcomes or multiple-time points or a combination of both. No other statistically significant effects were observed for other variables.

4.2. Assessment reliability studies

4.2.1. Study and sample characteristics

Fourteen individual studies published between the years of 1997 to 2019 met all inclusionary criteria and were included in the final analyses examining the reliability of outcomes associated with psychological and psychiatric assessments. Two studies were published in 1997 and three studies were published in 2019; a single study was published in all other years. A total of 83 individual outcomes were reported across studies and ultimately included in this meta-analysis. Methodological features and sample descriptive statistics of the 14 assessment reliability studies are presented in Tables 1 and 8, respectively. Descriptive statistics are reported for variables that were most consistently and clearly reported by study authors.

The final dataset included 332 examinees with a mean age of $M = 50.3$ years ($SD = 23.9$). The average number of assessment outcomes per study was 5.9 and the average number of participants assessed was 23.7. Nine studies (consisting of 73.5% of the outcomes analyzed) included both females and males, one study (12.0% of outcomes) included mostly males, one study (4.8% of outcome) included only males, and three studies (9.7% of outcomes) did not report participant gender. Four studies (55.4% of the outcomes included) were conducted in medical facilities, three (10.8%) in university clinics or the client's homes, two (8.4%) in other facilities (e.g., an undifferentiated hospital), one (4.8%) in a veterans hospital, and one (2.4%) in an inpatient facility. Three

Table 8
Sample descriptive statistics of included studies (Reliability Studies).

Study	Evaluator degree level	Number of evaluators	Were evaluators the same across modality?	Type of assessment	Mean client age	Client gender composition	Client race composition	Ave. number of clients evaluated	Number of outcomes per study
Chapman et al. (2019)	NA	1	Same	Neuropsychocological	65	Mixed	NA	24	1
Dekhtyar et al. (2019)	Masters	3	Same	Neuropsychological	55	Mixed	NA	10	3
Elford et al. (2000)	Doctoral; Medical	6	Different	Psychodiagnostic	9	Mixed	NA	23	10
Kobak, Williams, Jeglic, Salvucci, and Sharp (2008)	Masters; Doctoral	4	Different	Psychodiagnostic	NA	NA	NA	21	1
Montani et al. (1997)	Doctoral	2	Different	Psychodiagnostic	88	Mixed	NA	15	16
Porcari et al. (2009)	Bachelors; Masters; Doctoral; Medical	3	Different	Psychodiagnostic	NA	Male only	Mostly White/ Caucasian	20	4
Stain et al. (2011)	NA	NA	NA	Neuropsychological	20	Mixed	NA	11	4
Stead and Vinson (2019)	NA	NA	NA	Neuropsychological	NA	Mixed	NA	27	5
Temple et al. (2010)	Bachelors; Masters; Doctoral	NA	NA	Psychodiagnostic	39	NA	NA	19	4
Vahia et al. (2015)	NA	2	Same	Neuropsychological	71	Mixed	Multicultural	22	8
Wadsworth et al. (2016)	NA	NA	NA	Neuropsychological	65	Mixed	Native American	83	12
Wong et al. (2012)	Student	3	Different	Psychodiagnostic	42	NA	NA	28	3
Yoshino et al. (2001)	Medical	2	Different	Psychodiagnostic	49	Mostly Male	NA	14	10
Zarate et al. (1997)	Medical	NA	Different	Psychodiagnostic	NA	Mixed	NA	15	2

studies (18.2% of outcomes) did not report the type of facility where clients were assessed. Evaluations were conducted for psychodiagnostic purposes in eight studies (consisting of 60.2% of outcomes analyzed) and neuropsychological disorders were assessed in six studies (39.8% of outcomes). A random assignment procedure was used in four studies (including 42.2% of outcomes) and two studies (4.8% of outcomes) used a non-random, non-matched assignment procedure; the remaining eight studies (53.0% of the outcomes) did not report assignment procedures. No studies that met inclusionary criteria included the assessment of child or adolescent examinees.

4.2.2. *Three-level hierarchical linear model (HLM3) reliability assessment*

To properly capture the variance components at the client sampling-, outcome-, and study-levels and to account for interdependency between outcomes within studies, a three-level HLM as implemented in the R *metaSEM* (Cheung, 2015) package was also used in the analysis of assessment reliability outcomes, despite a smaller number of studies and outcomes within the included studies. Results of the HLM3 model revealed that the overall estimated effect size of the differences in reliability across the two assessment modalities using dependent samples was not statistically significant, Hedges' $g = 0.07$, 95% CIs [−0.02, 0.17], ($SE = 0.05$), $p = .160$. That is, no differences between the results or opinions generated from VCT-conducted evaluations and those generated from in-person evaluations were observed. The I^2_2 and I^2_3, were 0.81 and 0.09, respectively, demonstrating that most variations in the method of assessment (VCT versus in-person) can be attributed to the variation between outcomes.

5. Discussion

Advances in technology have virtually transformed the way people communicate and interact with each other. It is not surprising, then, that the mental health sector is embracing video communication systems in an effort to keep up with existing social norms and take advantage of the opportunity to reach clients who may otherwise have limited options for local specialty providers, face increased costs and lost wages for travel to distant clinics, or go without services altogether. Beyond this, remote technologies broadly have become a necessity in the wake of COVID-19 to sustain clinical practices and safely engage clients in need. The use of VCT for delivering a wide range of mental and behavioral health services is not just a growing practice—it is exploding across the country and the globe, with many looking to these technologies as a sort of savior in the fight to resolve the mental health crises by improving accessibility and affordability. Yet, it could be argued that the high degree of hope and promise placed in VCT approaches may be disproportionate to the existing literature base supporting its use. For example, while most research to date has demonstrated adequate comparability to traditional in-person mechanisms, there remains some concern among the professional community that VCT may present with its own unique set of barriers (e.g., the loss of relevant behavioral observations that either occur beyond camera view or obscured by the quality of the video, technological issues that negatively disrupt services, inability to administer certain types of testing instruments; Batastini et al., 2019; Simpson, 2001). Although professional practice guidelines exist, as well as various publications dedicated to enhancing the efficacy of VCT practices (Campbell, Millán, & Martin, 2018; Luxton, Nelson, & Maheu, 2016), additional research on the aggregate effects of VCT when directly compared to in-person is needed to vindicate the field's enthusiasm. While not the first study to compile existing research related to the use of VCT in conducting mental and behavioral health services, it is the first known meta-analysis to do so in a generalized manner that included treatment efficacy and assessment reliability outcomes across disciplines, clinical settings, and client populations. Furthermore, in taking a more holistic look at this literature, we aimed to uncover common gaps in how VCT practices are researched, and how findings are subsequently reported, that currently limit the totality of our understanding about VCT in general and in what nuanced ways it is distinguished (or not) from in-person services.

In this study, separate meta-analyses were conducted on mental health and behavioral outcomes associated with intervention implementation ($k = 43$ studies) and the reliability of assessments for making clinically relevant decisions ($k = 14$) that included a combined total of 4668 participants and 364 independent outcomes. In both sets of analyses, we expected to observe non-significant differences between VCT and in-person modalities, accompanied by small effect size estimates. Overall, findings supported this general hypothesis, such that VCT and in-person interventions produced similar changes from pre- to post-intervention (most of which focused on reductions in mental health symptomology using client self-report instruments) and assessments conducted via VCT yielded similar measurement scores as in-person assessments. For the intervention studies, these findings were supported using both a conventional meta-analytic approach and a three-level hierarchical linear model to account for nested variance. From the HLM3, it appeared that the small and non-significant observed differences between VCT and in-person interventions were largely explained by extraneous factors (i.e., the study itself, how/what outcomes were measured, and participant sampling) rather than differences associated with the delivery modality.

Interestingly, results of the conventional moderator analysis examining treatment effects across modalities revealed that, when the sample included mostly (i.e., over 80%) male clients, in-person interventions led to greater improvements than VCT, indicating a possibility that men benefit more from traditional in-person delivery formats. It should be noted, however, that studies including mostly male clients were predominantly from studies focused on veteran populations. Therefore, this effect may not generalize to other groups of male clientele. Furthermore, the more robust HLM3 moderator analysis failed to replicate an effect of the mostly males gender category. Both moderator analyses did, however, show that when a study's gender breakdown consisted of all or mostly female clients, more positive post-intervention changes were found with the use of VCT compared to in-person. Some evidence suggests that women are more frequent technology users than men, preferring to connect via text messaging, social media, and online video calls (Kimbrough, Guadagno, Muscanell, & Dill, 2013). Perhaps VCT appeals more to these preferences, leading to better rapport, engagement, and/or compliance. Notably, the four studies that had a stronger focus on women collectively included rural, Hispanic, medically underserved, veteran, and disordered eating clients. Therefore, the factors contributing to modality differences may not be isolated to gender.

Both the conventional analysis and HLM3 approach also consistently revealed that the intervention site may matter to some extent given that VCT produced better outcomes within medical settings than in-person interventions. While the reason behind this effect could not be further explored (largely due to the extent of unreported information within studies), it may be the case that medical facilities have the capacity to acquire more technologically advanced equipment than facilities tailored primarily to consumers of mental health services and/or that the availability of VCT in these settings opens a wider door to higher-quality providers and multidisciplinary treatment teams. Medical facilities may also inherently treat a certain subtype of clients who benefitted in some unique way from VCT care or the specific type of services delivered in this manner. For example, it is possible that these facilities end up serving more severe/higher-risk clients for whom the convenience of VCT may have increased compliance or willingness to engage. Consider a highly agoraphobic client who may feel more comfortable talking to a provider through the physical barrier of a screen. Or, if clients were more physically compromised, the use of technology may have offered needed accommodations (e.g., closed captioning, ability to increase volume) that cannot be modified in-person meetings. Importantly, based on reductions in heterogeneity in the HLM3 models, gender and site type differences may be more attributable to the specific study, suggesting that these effects may not replicate as well across studies. None of the comparisons between VCT and in-person interventions were affected by the research design or the primary clinical disorder treated.

Notably, there were three outliers in the intervention outcomes analysis. Gros et al. (2011) supported the use of an in-person exposure therapy intervention for veterans over VCT; however, further review of this individual study revealed that VCT was still associated with large pre-post effect sizes in reductions of self-reported PSTD symptoms, anxiety, depression, stress, and interference of symptoms. Similarly, in Kelleher et al. (2019), a behavioral VCT pain management intervention produced nearly identical improvements as the traditional in-person intervention. The outcomes reported in Moreno et al. (2012), on the other hand, favored the VCT condition. Specifically, a webcam-based telepsychiatry intervention for Hispanic clients lead to greater symptom reductions across measures of depression, quality of life, and functional ability compared to a treatment as usual condition. Thus, these outliers do not appear to disconfirm the utility of VCT.

5.1. Strengths, limitations, and gaps to fill

To broadly examine whether VCT is comparable to in-person services for improving mental and behavioral health outcomes, this study was strengthened by the application of two approaches to analyze the common effect size for outcomes nested within a study: (1) a conventional approach using manual adjustment to stochastically-dependent outcomes following the method proposed by Hedges and Olkin (1985) and (2) an HLM3 approach. Both methods, although to a certain extent duplicative, were necessary. The conventional approach allowed insights into the analysis of study outliers and case diagnostics. The HLM3 approach, on the other hand, allowed for a more efficient method of dealing with dependencies between outcomes grouped by their respective studies and did not require an explicit priori specification of the level of dependencies. The HLM3 approach also addressed the issue of portioning the total variance to within-outcome (participants sampling), as well as between-outcome and between-study variances, and provided further insights into the moderators' effects. The conventional approach, although more robust in terms of case diagnostics, required manual adjustments to deal with the stochastic dependencies between outcomes. It was assumed that the level of correlations of nested outcomes was 0.5; however, the exact correlations between outcomes were not known. In reality, this level of dependency might have been too aggressive for less related outcomes (e.g., psychosocial functioning and care needs) or too liberal for more related outcomes or those being observed repeatedly (e.g., the same subjects repeatedly evaluated on depression levels in 3-month increments). Unless future studies routinely begin reporting correlations between outcome variables, this limitation will be difficult to overcome.

An HLM3 approach was also used for the reliability of psychological and psychiatric evaluations, where the focus was to examine whether the mechanism of assessment (using VCT or in-person) produced a statistically significant difference in the observed outcomes, typically related to a diagnostic or treatment decision. Using this approach, reported statistics, means and corresponding SDs, t. χ^2 parameters, and the corresponding n or p-values could be used to convert all outcomes into Hedges' g effect sizes using formulas for within-subjects designs. To be consistent with the conventional analysis used in the between-studies analysis of intervention outcomes, a 0.5 correlation coefficient was likewise used. Although a similar logic in selecting 0.5 could be applied, the use of this coefficient was a more arbitrary choice in this case. Arguably, however, the greatest shortcoming of the reliability analysis was the small number of studies that met inclusionary criteria and the relatively small average number of outcomes within each of these studies ($k = 6.4$).

While these meta-analyses drew strengths from both its inclusivity of mental health related services and the multi-method analytical approach, a meta-analysis is only as powerful as the literature base on which it is derived. Across both types of studies—intervention outcome and assessment reliability—there were several commonly occurring issues that caused unresolvable barriers to the present investigation, and therefore, inherently limit the meta-analytic results. First, research designs varied widely (e.g., single condition with descriptive statistics, within-subjects, between-subjects, mixed-designs) such that it was difficult to find enough studies that could be analyzed together. And, in other studies we attempted to include, the research design was too niche to meta-analyze at all. One study, for example, compared one in-person group treatment for smoking cessation with six different VCT groups, using smoking "quit rates" as the only outcome measure of interest (Carlson et al., 2011). The use of nested groups, or the existence of several types of remote service groups within one condition, coupled with the inclusion of dependent measures that could not easily or logically be combined with more diagnostically driven measures prevented this study from being included.

Second, few studies actually compared VCT to an in-person comparison group, limiting the number of studies that could be included in this review. Of the 125 full-text articles retained for double-coding, 27 (21.6%) were dropped because they used a pre-post treatment design only. Twelve coded studies (9.6%) did not report the method of group comparisons, and such information was unable to be gleaned from what was reported. Many studies that were collected also only focused on the therapeutic process or surveyed service satisfaction but did not measure whether the service achieved its intended objectives. Further, studies that did meet inclusionary criteria were based on relatively small sample sizes, ranging from 16 to 470 intervention clients and 4 to 84 assessment cases. To have greater confidence in the use of VCT for mental healthcare purposes, it is important to know whether VCT is similarly effective to how services have traditionally been provided, not just whether it is better than nothing. More research is clearly needed on direct comparisons between VCT and in-person modalities.

Third, the generalizability of findings is limited by the fact that a larger proportion of participants across studies were adult, white, male, and/or treated through the VA ($k = 12$). Though statistically significant differences did not emerge with respect to VA settings, the VA has been at the forefront of VCT implementation and, as a result, may have developed better ways of adapting technology that minimize any differences between service modalities. Therefore, empirical efforts are needed to examine VCT use with different demographic groups and in a wider range of clinical settings. Fourth, and relatedly, all included studies were available in the English language. It is possible that international studies unavailable in an English translation may have offered meaningful insights.

Finally, and perhaps most troubling, the process of collecting and coding studies for this meta-analysis revealed the inconsistent and, in more cases than expected, overly minimalist approach to reporting the study's methodology and results. For example, although most between-subjects studies used a random assignment procedure, some neglected to report the method of participant assignment. We were additionally surprised by the frequency with which studies failed to report basic participant demographics such as ethnicity or age, the specific technologies used to deliver services (39.3% of studies did not), or sufficient test statistics (e.g., a few studies included charts or graphs that did not display or otherwise report exact values). Most attempts to contact corresponding authors were not fruitful. Studies were also generally lacking in their description of services provided (e.g., describing a service simply as a "telepsychiatry ambulatory clinic," Modai et al., 2006; "outpatient psychiatric care," Farabee, 2016; or including "pertinent" CBT sessions, De Las Cuevas et al., 2006), whether any noteworthy differences in service delivery or content occurred as a byproduct of the modality, or the steps taken to ensure treatment equivalency (e.g. measures of provider fidelity). For purposes of this analysis, it was assumed that intervention and assessment techniques were generally equivalent save for the mode of delivery.

Due to missing or unclear reporting, we were unable to code many of the variables intended, thereby limiting the rigorousness of our moderator analyses. For example, it would have been especially meaningful to test whether certain technology systems or bandwidths produced better outcomes. Additionally, our ability to assess the risk of bias (as is customary in meta-analyses) posed by study parameters and design features (e.g., concealment of participant allocation procedures, blinding of participants or personnel, treatment implementation; see Cochrane's GRADE handbook[4]) was not practical due to poor reporting, as well as the prevalence of studies that were conducted in more real-world contexts with fewer internal controls (or assumed to have few controls in the absence of any information to the contrary). We, therefore, cannot be certain these types of biases (see Munder & Barth, 2018) did not impact or explain the observed outcomes. However, as reported in Munder and Barth (2018), self-report symptoms (the primary approach to assessing outcomes among our included studies) do not appear to over-estimate treatment effects (see pg. 351). We also did not find evidence of differential outcomes based on general research design (random vs. non-random). Of course, there are myriad other ways in which bias may be introduced and compromise the validity of meta-analytic interpretations.

Based on these frustrations, we next offer a set of general guidelines to help standardize the scientific reporting of studies that evaluate the effectiveness or reliability of VCT and other remote technologies compared to in-person services.

5.2. Recommended guidelines for the reporting of Telemental Health Research

In addition to adhering to existing guidelines on designing and conducting efficacy studies related telemedicine practices (Krupinski & Bernard, 2014), including the evaluation of telemental health programs (Kramer et al., 2012), and generally accepted manuscript reporting criteria (See CONSORT, 2010), we strongly advise future researchers to explicitly report further details on the following:

1. ***Participants and/or target population of the service***
   - Who are the participants or who is receiving the service? For example, study participants could be patients or clients, or they could be caregivers, teachers, service providers, or multiple types of people. If the actual study participants were not patients/clients directly, were the participants receiving services on behalf of someone else (e.g., parents receiving training to manage their children with ADHD)?
   - Total sample size and sample size by condition (VCT vs. in-person).
   - Basic demographics (e.g., age, gender, diagnostic breakdown, ethnicity, educational level) for the total sample and by the condition. Also consider including population-specific demographic variables such as military branch for veterans, active duty service members, or the number of violent vs. non-violent index offenses for justice-involved persons.
   - Specific referral or inclusionary criteria (e.g., meeting criteria for a specific DSM-5 disorder, low cognitive functioning, suicide risk, substance misuse, caring for a parent with dementia).
2. ***Site descriptors***
   - The site or location where VCT clients go to receive services (e.g., outpatient clinic, general medical facility, VA hospital, correctional institution, academic setting, their homes). Of note, it is not always enough to report the name of the facility or institution where services were rendered, as it is often not clear what type of

[4] https://gdt.gradepro.org/app/handbook/handbook.html

site it is or the level of patient/client care management (e.g., inpatient, outpatient, closed custody), and this information is sometimes difficult to discern or not readily available by other means (e.g., internet searches).
- The site or location where in-person clients go to receive services.
- The site or location where providers deliver services.

3. ***Technology descriptors***
   - Software programs used (e.g., Skype, Polycom, Cisco, Microsoft Lync).
   - Equipment setup (i.e., what type of equipment did providers vs. clients use?).
   - Equipment access (e.g., was equipment provided to clients by the site or were clients required to use their own?).
   - Connection type and speed and whether these differed for clients vs. providers.
   - Video resolution (e.g., 710p, 1080i, 1080p) and whether this differed for clients vs. providers.
   - Frequency of technological problems (e.g., related to video connection loss, audio delay, picture distortions, security or confidentiality breaches, logistical issues encountered).
4. ***Intervention descriptors***
   - Assessment and screening procedures for inclusion in the intervention (e.g., was an in-person intake required before using VCT or were all aspects of the VCT condition completed remotely?).
   - Primary goal(s) of the intervention (e.g., symptom reduction, medication compliance, behavioral management, increase interpersonal skills).
   - Type of service(s) provided (e.g., counseling/therapy, psychoeducation or training, psychotropic medication management, psychodiagnostic assessment).
   - Structure of services (e.g., number of sessions or contacts with clients, length of sessions, primary theoretical orientation or framework applied, whether services were provided individually or in a group setting).
   - Differences in service type or structure across modalities (e.g., were any modifications required to uniquely accommodate a virtual platform?).
5. ***Provider descriptors***
   - The number of providers involved and whether they only provided services via one delivery mechanism or multiple (e.g., was one provider assigned to VCT and one to in-person, did one provider deliver all services?).
   - Any gender, age, ethnicity, or educational differences in VCT vs. in-person providers. Basic demographics are perhaps especially relevant when providers are a primary focus of the study, such as in reliability research, as client-therapist variables may contribute to meaningful differences in assessment-related outcomes.
   - Any specialized training requirements for providers regarding the use of VCT or other remote technologies (e.g., was on-site training provided, were a certain number of continuing education credits required?).
6. ***Research design descriptors***
   - Use of group equivalence procedures.
   - Whether the study was within-subjects, between-subjects, or mixed.
   - The number of data collection timepoints or follow-up assessments.
   - Final sample size used in each set of analyses and attrition rates at all phases of data collection.

While this is not an exhaustive list, as studies on specific samples or for more specialized purposes may warrant the reporting of additional data, the recommendations provided above bore from information that was commonly missing from or reported in a manner that was confusing or incomplete in studies gathered for this meta-analysis. That is, these recommendations should be considered fundamental and at-a-minimum. Given the number of studies without sufficient statistics, we also remind researchers to report all statistical procedures used, along with relevant test statistics, means, standard deviations, effect size estimates, and confidence intervals for all comparisons. As noted earlier, it would also be helpful for future meta-analytical research if correlations between outcomes were calculated and reported; otherwise, aggregate effects could be inflated if these correlations are not properly controlled. Again, researchers are encouraged to review the CONSORT statement for more generic standards for reporting treatment outcome studies; adherence to these standards will ensure that future meta-analyses can better assess aggregate outcomes, moderators, and bias risks.

6. Final conclusions

There is no question that remote delivery mechanisms are hitting a stride in mental and behavioral health service industries. But, are VCT services just as good as those delivered in-person? Thus far, the answer points to “yes,” as the available evidence suggests VCT does not grossly impede clinical outcomes in an overall general sense. Yet, there is more work to be done. Stronger research designs, greater inclusivity regarding client demographics and service settings, purposeful examination of moderator variables, and more comprehensive reporting of study methods and key findings are needed. Not only must we improve the scientific quality of general efficacy studies (thereby improving confidence in our answer), but it is also time to start diving deeper into what works for whom and under what circumstances. We also look forward to additional meta-analyses aggregating the effects of other innovative delivery approaches, such as the use of Smartphone mobile applications and virtual reality.

Role of funding sources

This project was supported by the Federal Office of Rural Health Policy (FORHP), Health resources and Services Administration (HRSA), U.S. Department of Health and Human Services (HHS) under cooperative agreement award number 6 U66RH31459–02-03. Dr. Paprzycki was also partially supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number 5 U54 GM115428. The information, conclusions, and opinions expressed are those of the authors and no endorsement by FORHP, HRSA, HHS, or NIH is intended or should be inferred.

Contributors

Ashley B. Batastini lead, supervised, and designed the study; to include development of the codebook, training and supervising coders, and overseeing reliability procedures and primary data analysis. Dr. Batastini also wrote significant portions of the manuscript and served as primary editor of all sections of the manuscript. Peter Paprzycki was the primary statistician on this project; he conducted the majority of statistical analyses and wrote all respective sections of the manuscript pertaining to the description of the statistical plan and results. Of note, the Dr. Paprzycki’s work was done in affiliation with the University of Southern Mississippi; he is now affiliated with the University of Toledo. Ashley Jones assisted in conducting literature searches, coding articles, coordinating other research assistants, organizing and managing data, creating tables and references. Ms. Jones also wrote portions of the manuscript describing the study and sample demographics. Nina MacLean assisted in article coding and effect size extraction, organizing tables. Dr. MacLean wrote a majority of the introduction section. All authors have proofread the manuscript in its entirety and have approved the draft in its current form.

Funding

This project was supported by the Federal Office of Rural Health Policy (FORHP), Health resources and Services Administration (HRSA), U.S. Department of Health and Human Services (HHS) under cooperative agreement award number 6 U66RH31459–02-03. Dr. Paprzycki was also partially supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number 5 U54 GM115428. The information, conclusions, and opinions expressed are those of the authors and no endorsement by FORHP, HRSA, HHS, or NIH is intended or should be inferred.

Declaration of Competing Interest

None.

Acknowledgements

We would like to thank the following individuals for their assistance with coding, data extraction, and manuscript preparation: Christopher M. King, Ph.D., J.D., Stephanie Van-Horn, Ph.D., Lauren Coaker, M.S., Michael Lester, M.S., Alexandra Repke, M.A., Alexandra Teller, M.A., Riley Davis, M.A., Nathan Ross, Ph.D., Ja'Nisha Robinson, M.S., Madison Pike, B.A., Kendall Klumpp, B.A., Abigail Armstrong, B.A., and Olivia Miller, M.S.

References

Abel, E. A., Glover, J., Brandt, C. A., & Godleski, L. (2017). Recommendations for the reporting of telemental health (TMH) literature based on a systematic review of clinical video teleconferencing (CVT) and depression. *Journal of Technology in Behavioral Science, 2*(1), 28–40. https://doi.org/10.1007/s41347-016-0003-1.

Acierno, R., Gros, D. F., Ruggiero, K. J., Hernandez-Tejada, M. A., Knapp, R. G., Lejuez, C. W., … Tuerk, P. W. (2016). Behavioral activation and therapeutic exposure for post traumatic exposure for posttraumatic stress disorder: A noninferiority trial of treatment delivered in person versus home-based telehealth. *Depression and Anxiety, 33*(5), 415–423. https://doi.org/10.1002/da.22476.

American Psychiatric Association. (2018). Annual Meeting | AJMC. (n.d.). Retrieved February 25, 2020, from https://www.ajmc.com/conferences/apa-2018.

American Psychological Association. (2013). Guidelines for psychological evaluations in child protection matters. *The American Psychologist, 68*(1), 20. https://doi.org/10.1037/a0029891.

American Telemedicine Association. (2013). *Practice guidelines for video-based online mental Health services*. Washington, DC: American Telemedicine Association. https://doi.org/10.1089/tmj.2013.9989.

Antonacci, D. J., Bloch, R. M., Saeed, S. A., Yildirim, Y., & Talley, J. (2008). Empirical evidence on the use and effectiveness of telepsychiatry via videoconferencing: Implications for forensic and correctional psychiatry. *Behavioral Sciences & the Law, 26*(3), 253–269. https://doi.org/10.1002/bsl.812.

Assink, M., & Wibbelink, C. J. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology, 12*(3), 154–174.

Australia''s National Digital Health Strategy. (2018). Safe, seamless and secure: Australia's national digital health strategy. https://conversation.digitalhealth.gov.au/sites/default/files/adha-strategy-doc-2ndaug_0_1.pdf.

Ax, R. K., Fagan, T. J., Magaletta, P. R., Morgan, R. D., Nussbaum, D., & White, T. W. (2007). Innovations in correctional assessment and treatment. *Criminal Justice and Behavior, 34*(7), 893–905. https://doi.org/10.1177/0093854807301555.

Backhaus, A., Agha, Z., Maglione, M. L., Repp, A., Ross, B., Zuest, D., … Thorp, S. R. (2012). Videoconferencing psychotherapy: A systematic review. *Psychological Services, 9*(2), 111–131. https://doi.org/10.1037/a0027924.

Batastini, A. B., McDonald, B. R., & Morgan, R. D. (2013). *Videoteleconferencing in forensic and correctional practice. Telemental health: Clinical, technical, and administrative foundations for evidence-based practice* (pp. 251–271).

Batastini, A. B., & Morgan, R. D. (2016). Connecting the disconnected: Preliminary results and lessons learned from a telepsychology initiative with special management inmates. *Psychological Services, 13*(3), 283–291. https://doi.org/10.1037/ser0000078.

Batastini, A. B., Pike, M., Thoen, M. A., Jones, A. C., Davis, R. M., & Escalera, E. (2019). Perceptions and use of videoconferencing in forensic mental health assessments: A survey of evaluators and legal personnel. *Psychology, Crime & Law*, 1–21. https://doi.org/10.1080/1068316X.2019.1708355.

Benavides-Vaello, S., Strode, A., & Sheeran, B. C. (2013). Using technology in the delivery of mental health and substance abuse treatment in rural communities: A review. *The Journal of Behavioral Health Services & Research, 40*(1), 111–120. https://doi.org/10.1007/s11414-012-9299-6.

Blevins, C. A., Weathers, F. W., Davis, M. T., Witte, T. K., & Domino, J. L. (2015). The posttraumatic stress disorder checklist for DSM-5 (PCL-5): Development and initial psychometric evaluation. *Journal of Traumatic Stress, 28*, 489–498. https://doi.org/10.1002/jts.22059.

Borenstein, M. (2009). Effect sizes for continuous data. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 221–235). Russell Sage Foundation.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2014). *Comprehensive meta-analysis (version 3.3.070) [computer software]*. Englewood, NJ: Biostat.

Bouchard, S., Paquin, B., Payeur, R., Allard, M., Renaud, P., & Lapierre, J. (2004). Delivering cognitive-behavior therapy for panic disorder with agoraphobia in videoconference. *Telemedicine Journal and E-Health, 10*(1), 13–25. https://doi.org/10.1089/153056204773644535.

Bovin, M. J., Marx, B. P., Weathers, F. W., Gallagher, M. W., Rodriguez, P., Schnurr, P. P., & Keane, T. M. (2016). Psychometric properties of the PTSD checklist for diagnostic and statistical manual of mental disorders-fifth edition (PCL-5) in veterans. *Psychological Assessment, 28*(11), 1379–1391. https://doi.org/10.1037/pas0000254.

Brearly, T. W., Shura, R. D., Martindale, S. L., Lazowski, R. A., Luxton, D. D., Shenal, B. V., & Rowland, J. A. (2017). Neuropsychological test administration by videoconference: A systematic review and meta-analysis. *Neuropsychology Review, 27*(2), 174–186. https://doi.org/10.1007/s11065-017-9349-1.

Campbell, L. F., Millán, F. A., & Martin, J. N. (2018). A telepsychology casebook: Using technology ethically and effectively in your professional practice. *American Psychological Association.*. https://doi.org/10.1037/0000046-000.

Carlson, L. E., Lounsberry, J. J., Maciejewski, O., Wright, K., Collacutt, V., & Taenzer, P. (2011). Telehealth-delivered group smoking cessation for rural and urban participantsL feasibility and cessation rates. *Addictive Behaviors, 37*, 108–114. https://doi.org/10.1016/j.addbeh.2011.09.011.

Centers for Medicare and Medicaid Services. (2020). Medicare telemedicine health care provider fact sheet. May. Retrieved from https://www.cms.gov/newsroom/fact-sheets/medicare-telemedicine-health-care-provider-fact-sheet.

Chapman, J. E., Cadilhac, D. A., Gardner, B., Ponsford, J., Bhalla, R., & Stolwyk, R. J. (2019). Comparing face-to-face and videoconference completion of the Montreal Cognitive Assessment (MoCA) in community-based survivors of stroke. *Journal of Telemedicine and Telecare*, 1–9. https://doi.org/10.1177/1357633X19890788.

Cheung, M. (2019a). A guide to conducting a meta-analysis with non-independent effect sizes. *Neuropsychology Review, 29*(4), 387–396. https://doi.org/10.1007/s11065-019-094156.

Cheung, M. W. (2019b). *Meta-analysis: A structural equation modeling approach.* Chichester, West Sussex: John Wiley & Sons.

Cheung, M. W. L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods, 19*(2), 211–229. https://doi.org/10.1037/a0032968.

Cheung, M. W. L. (2015). metaSEM: An R package for meta-analysis using structural equation modeling. *Frontiers in Psychology, 5*, 1521. https://doi.org/10.3389/fpsyg.2014.01521.

Choi, N. G., Marti, C. N., Bruce, M. L., Hegel, M. T., Wilson, N. L., & Kunik, M. E. (2014). Six-month postintervention depression and disability outcomes of in-home telehealth problem-solving therapy for depressed, low-income homebound older *adults*. *Depression and Anxiety, 31*(8), 653–661. https://doi.org/10.1002/da.22242.

Comer, J. S., Furr, J. M., Kerns, C. E., Miguel, E., Coxe, S., Elkins, R. M., … Freeman, J. B. (2017). Internet-delivered, family-based treatment for early-onset OCD: A pilot randomized trial. *Journal of Consulting and Clinical Psychology, 85*(2), 178–186. https://doi.org/10.1037/ccp0000155.

Comer, J. S., Furr, J. M., Miguel, E. M., Cooper-Vince, C. E., Carpenter, A. L., Elkins, R. M., … Chase, R. (2017). Remotely delivering real-time parent training to the home: An initial randomized trial of internet-delivered Parent-Child Interaction Therapy (I-PCIT). *Journal of Consulting and Clinical Psychology, 85*(9), 909–917. https://doi.org/10.1037/ccp0000230.

Day, S. X., & Schneider, P. L. (2002). Psychotherapy using distance technology: A comparison of face-to-face, video, and audio treatment. *Journal of Counseling Psychology, 49*(4), 499–503. https://doi.org/10.1037/0022-0167.49.4.499.

De la Cuevas, C., Arrendondo, T., Cabrera, M. F., Sulzenbacher, H., & Meise, U. (2006). Randomized clinical trial of telepsychiatry through videoconference versus face-to-face conventional psychiatric treatment. *Telemedicine and e-health, 12*(3), 341–350. https://doi.org/10.1089/tmj.2006.12.341.

Dekhtyar, M., Braun, E. J., Billot, A., Foo, L., & Kiran, S. (2019). Videoconference administration of the Western aphasia battery-revised: Feasibility and validity. *American Journal of Speech-Language Pathology*, 1–15.

Digital Health & Care Scotland. (2018). Scotland's digital health and care strategy: Enabling, connecting and empowering. , April 25 https://www.gov.scot/publications/scotlands-digital-health-care-strategy-enabling-connecting-empowering/.

Drago, A., Winding, T. N., & Antypa, N. (2016). Videoconferencing in psychiatry, a meta-analysis of assessment and treatment. *European Psychiatry, 36*, 29–37. https://doi.org/10.1016/j.eurpsy.2016.03.007.

Elford, R., White, H., Bowering, R., Ghandi, A., Maddiggan, B., St. John, K., … Battcock, A. (2000). A randomized, controlled trial of child psychiatric assessments conducted using videoconferencing. *Journal of Telemedicine and Telecare, 6*, 73–82. https://doi.org/10.1258/1357633001935086.

Farabee, D., Calhoun, S., & Veliz, R. (2016). An experimental comparison of telepsychiatry and conventional psychiatry for parolees. *Psychiatric Services, 67*(5), 562–565. https://doi.org/10.1176/appi.ps.201500025.

Fortney, J. C., Maciejewski, M. L., Tripathi, S. P., Deen, T. L., & Pyne, J. M. (2011). A budget impact analysis of telemedicine-based collaborative care for depression. *Medical Care, 49*(9), 872–880. https://doi.org/10.1097/MLR.0b013e31821d2b35.

Fortney, J. C., Pyne, J. M., Edlund, M. J., Williams, D. K., Robinson, D. E., Dinesh, M., & Henderson, K. L. (2007). A randomized trial of telemedicine-based collaborative care for depression. *Journal of General Internal Medicine*, 1086–1093. https://doi.org/10.1007/s11606-007-0201-9.

Fortney, J. C., Pyne, J. M., Mouden, S. B., Mittal, D., Hudson, T. J., Schroeder, G. W., … Rost, K. M. (2013). Practice-bases versus telemedicine-based collaborative care for depression in rural federally qualified health centers: A pragmatic randomized comparative effectiveness trial. *The American Journal of Psychiatry, 170*, 414–425. https://doi.org/10.1176/appi.ajp.2012.12050696.

Frueh, B. C. (2015). Solving mental healthcare access problems in the twenty-first century. *Australian Psychologist, 50*(4), 304–306. https://doi.org/10.1111/ap.12140.

Frueh, B. C., Monnier, J., Yim, E., Grubaugh, A. L., Hamner, M. B., & Knapp, R. G. (2007). A randomized trial of telepsychiatry for pot-traumatic stress disorder. *Journal of Telemedicine and Telecare, 13*, 142–147. https://doi.org/10.1258/135763307780677604.

Germain, V., Marchand, A., Bouchard, S., Drouin, M. S., & Guay, S. (2009). Effectiveness of cognitive behavioural therapy administered by videoconference for posttraumatic stress disorder. *Cognitive Behaviour Therapy, 38*(1), 42–53. https://doi.org/10.1080/16506070802473494.

Germain, V., Marchand, A., Bouchard, S., Guay, S., & Drouin, M. S. (2010). Assessment of the therapeutic alliance in face-to-face or videoconference treatment for posttraumatic stress disorder. *Cyberpsychology, Behavior and Social Networking, 13*, 29–35. https://doi.org/10.1089/cyber.2009.0139.

Gershkovich, M., Herbert, J. D., Forman, E. M., & Glassman, L. (2016). Guided internet-based self-help intervention for social anxiety disorder with videoconferenced therapist support. *Cognitive and Behavioral Practice, 23*(2), 239–255. https://doi.org/10.1016/j.cbpra.2015.08.005.

Glassman, L. H., Mackintosh, M., Talkovsky, A., Wells, S. Y., Walter, K. H., Wickramasinghe, I., & Morland, L. A. (2019). Quality of life following treatment for PTSD: Comparison of videoconferencing and in-person modalities. *Journal of Telemedicine and Telecare, 25*(2), 123–127. https://doi.org/10.1177/1357633X17740610.

Glynn, L. H., Chen, J. A., Dawson, T. C., Gelman, H., & Zeliadt, S. B. (2020). Bringing chronic-pain care to rural veterans: A telehealth pilot program description. *Psychological Services*. https://doi.org/10.1037/ser0000408. Advance online publication.

Grob, P., Weintraub, D., Sayles, D., Raskin, A., & Ruskin, P. (2001). Psychiatric assessment of a nursing home population using audiovisual telecommunication. *Journal of Geriatric Psychiatry and Neurology, 14*(2), 63–65. https://doi.org/10.1177/089198870101400203.

Gros, D. F., Yoder, M., Tuerk, P. W., Lozano, B. E., & Acierno, R. (2011). Exposure therapy for PTSD delivered to veterans via telehealth: Predictors of treatment completion and outcome and comparison to treatment delivered in person. *Behavior Therapy, 42*, 276–283. https://doi.org/10.1016/j.beth.2010.07.005.

Halphen, J. M., Dyer, C. B., Lee, J. L., Reyes-Ortiz, C. A., Murdock, C. C., Hiner, J. A., & Burnett, J. (2020). Capacity evaluations for adult protective services: Videoconference or in-person interviews. *Journal of Elder Abuse & Neglect, 32*(2), 121–133. https://doi.org/10.1080/08946566.2020.1740127.

Hedges, L., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*(2), 107–128. https://doi.org/10.3102/10769986006002107.

Herbert, M. S., Afari, N., Liu, L., Heppner, P., Rutledge, T., Williams, K., … Wetherell, J. L. (2017). Telehealth versus in-person acceptance and commitment therapy for chronic pain: A randomized noninferiority trial. *The Journal of Pain, 18* (2), 200–211. https://doi.org/10.1016/j.jpain.2016.10.014.

Hilty, D. M., Ferrer, D. C., Parish, M. B., Johnston, B., Callahan, E. J., & Yellowlees, P. M. (2013). The effectiveness of telemental health: A 2013 review. *Telemedicine Journal and E-Health, 19*(6), 444–454. https://doi.org/10.1089/tmj.2013.0075.

Himle, M. B., Freitag, M., Walther, M., Franklin, S. A., Ely, L., & Woods, D. W. (2012). A randomized pilot trial comparing videoconference versus face-to-face delivery of behavior therapy for childhood tic disorders. *Behaviour Research and Therapy, 50*, 565–570. https://doi.org/10.1016/j.brat.2012.05.009.

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications*. New York, NY: Routledge.

Hulsbosch, A. M., Nugter, M. A., Tamis, P., & Kroon, H. (2017). Videoconferencing in a mental health service in The Netherlands: A randomized controlled trial on patient satisfaction and clinical outcomes for outpatients with severe mental illness. *Journal of Telemedicine and Telecare, 23*(5), 513–520. https://doi.org/10.1177/1357633x16650096.

Hungerbuehler, I., Valiengo, L., Loch, A. A., Rössler, W., & Gattaz, W. F. (2016). Home-based psychiatric outpatient care through videoconferencing for depression: A randomized controlled follow-up trial. *JMIR Mental Health, 3*(3), Article e36. https://doi.org/10.2196/mental.5675.

Hunter, J., & Schmidt, F. (2015). *Methods of meta-analysis : Correcting error and bias in research findings* (3rd ed.). Thousand Oaks, CA: Sage.

Hyler, S. E., Gangure, D. P., & Batchelder, S. T. (2005). Can telepsychiatry replace inperson psychiatric assessments? A review and meta-analysis of comparison studies. *CNS Spectrums, 10*(5), 403–415. https://doi.org/10.1017/S109285290002277X.

Kelleher, S. A., Winger, J. G., Dorfman, C. S., Ingle, K. K., Moskovich, A. A., Abernethy, A. P., … Somers, T. J. (2019). A behavioral cancer pain intervention: A randomized noninferiority trial comparing in-person with videoconference delivery. *Psycho-Oncology, 28*, 1671–1678. https://doi.org/10.1002/pon.5141.

Kimbrough, A. M., Guadagno, R. E., Muscanell, N. L., & Dill, J. (2013). Gender differences in mediated communication: Women connect more than do men. *Computers in Human Behavior, 29*(3), 896–900. https://doi.org/10.1016/j.chb.2012.12.005.

King, V. L., Brooner, R. K., Peirce, J. M., Kolodner, K., & Kidorf, M. S. (2014). A randomized trial of web-based videoconferencing for substance abuse counseling. *Journal of Substance Abuse Treatment, 46*(1), 36–42. https://doi.org/10.1016/j.jsat.2013.08.009.

Kobak, K. A., Williams, J. B. W., & Engelhardt, N. (2008). A comparison of face-to-face and remote assessment of inter-rater reliability on the Hamilton depression rating scale via videoconferencing. *Psychiatry Research, 158*(1), 99–103. https://doi.org/10.1016/j.psychres.2007.06.025.

Kobak, K. A., Williams, J. B. W., Jeglic, E., Salvucci, D., & Sharp, I. R. (2008). Face-to-face versus remotes administration of the Montgomery-asberg depression rating scale using videoconferencing and telephone. *Depression and Anxiety, 25*, 913–919. https://doi.org/10.1002/da.20392.

Kramer, G. M., Shore, J. H., Mishkind, M. C., Friedl, K. E., Poropatich, R. K., & Gahm, G. A. (2012). A standard telemental health evaluation model: The time is now. *Telemed. J. e-Health, 18*, 309–313. https://doi.org/10.1089/tmj.2011.0149.

Krupinski, E. A., & Bernard, J. (2014, March). Standards and guidelines in telemedicine and telehealth. In *Healthcare (Vol. 2, No. 1, pp. 74-93)*. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/healthcare2010074.

Langarizadeh, M., Tabatabaei, M. S., Tavakol, K., Naghipour, M., Rostami, A., & Moghbeli, F. (2017). Telemental health care, an effective alternative to conventional mental care: A systematic review. *Acta Informatica Medica, 25*(4), 240. https://doi.org/10.5455/aim.2017.25.240-246.

LaRock, Z. (2019). *Telemental health report: How telehealth can help US hospitals and health systems manage the $238 billion mental health crisis*. Business Insider.

Larson, J. L., Rosen, A. B., & Wilson, F. A. (2018). The effect of telehealth interventions on quality of life of cancer patients: A systematic review and meta-analysis. *Telemedicine and e-Health, 24*(6), 397–405. https://doi.org/10.1089/tmj.2017.0112.

Lerman, A. F., Kim, D., Ozinal, F. R., & Thompson, T. E. (2018). Telemental and Telebehavioral Health Considerations: A 50-State Analysis on the Development of Telehealth Policy. In *Telehealth and Medicine Today*.

Liu, L., Thorp, S. R., Moreno, L., Wells, S. Y., Glassman, L. H., Busch, A. C., Zamora, T., Rodgers, C. S., Allard, C. B., Morland, L. A., & Agha, Z. (2019). Videoconferencing psychotherapy for veterans with PTSD: Results from a randomized controlled non-inferiority trial. Journal of Telemedicine and Telecare, *0*(0), 1–13. doi:https://doi.org/10.1177/1357633X19853947.

Liu, S., Yang, L., Zhang, C., Xiang, Y. T., Liu, Z., Hu, S., & Zhang, B. (2020). Online mental health services in China during the COVID-19 outbreak. *Lancet Psychiatry, 7*(4), e17–e18. https://doi.org/10.1016/S2215-0366(20)30077-8.

Luxton, D. D., Nelson, E. L., & Maheu, M. M. (2016). A practitioner's guide to telemental health: How to conduct legal, ethical, and evidence-based telepractice. *American Psychological Association.*. https://doi.org/10.1037/14938-000.

Luxton, D. D., Pruitt, L. D., Wagner, A., Smolenski, D. J., Jenkins-Guarnieri, M. A., & Gahm, G. (2016). Home-based telebehavioral health for U.S. military personnel and veterans with depression: A randomized controlled trial. *Journal of Consulting and Clinical Psychology, 84*(11), 923–934. https://doi.org/10.1037/ccp0000135.

Maieritsch, K. P., Smith, T. L., Hessinger, J. D., Ahearn, E. P., Eickhoff, J. C., & Zhao, Q. (2016). Randomized controlled equivalence trial comparing videoconference and in person delivery of cognitive processing therapy for PTSD. *Journal of Telemedicine and Telecare, 22*(4), 238–243. https://doi.org/10.1177/1357633x15596109.

Manchanda, M., & McLaren, P. (1998). Cognitive behaviour therapy via interactive video. *Journal of Telemedicine and Telecare, 4*(1), 53–55. https://doi.org/10.1258/1357633981931452.

Matthews, K. (2017). Telepsychiatry: The modern solution to our mental health. In *TechTree*. August. Retrieved at https://www.techtree.com/content/features/13159/telepsychiatry-modern-solution-mental-health.html.

Miller, T. W., Clark, J., Veltkamp, L. J., Burton, D. C., & Swope, M. (2008). Teleconferencing model for forensic consultation, court testimony, and continuing education. *Behavioral Sciences & the Law, 26*(3), 301–313. https://doi.org/10.1002/bsl.809.

Mitchell, J. E., Crosby, R. D., Wonderlich, S. A., Crow, S., Lancaster, K., Simonich, H., … Myers, T. C. (2008). A randomized trial comparing the efficacy of cognitive-behavioral therapy for bulimia nervosa delivered via telemedicine versus face-to-face. *Behaviour Research and Therapy, 46*, 581–592. https://doi.org/10.1016/j.brat.2008.02.004.

Modai, I., Jabarin, M., Kurs, R., Barak, P., Hanan, I., & Kitain, L. (2006). Cost effectiveness, safety, and satisfaction with video telepsychiatry versus face-to-face care in ambulatory settings. *Telemedicine journal and e-health: the official journal of the American Telemedicine Association, 12*, 515–520. https://doi.org/10.1089/tmj.2006.12.515.

Montani, C., Billaud, N., Tyrrell, J., Fluchaire, I., Malterre, C., Lauvernay, N., … Franco, A. (1997). Psychological impact of a remote psychometric consultation with hospitalized elderly people. *Journal of Telemedicine and Telecare, 3*, 140–145. https://doi.org/10.1258/1357633971931048.

Moreno, F. A., Chong, J., Dumbauld, J., Humke, M., & Byreddy, S. (2012). Use of standard webcam and internet equipment for telepsychiatry treatment of depression among underserved Hispanics. *Psychiatric Services, 63*(12), 1213–1217. https://doi.org/10.1176/appi.ps.201100274.

Morgan, R. D., Patrick, A. R., & Magaletta, P. R. (2008). Does the use of telemental health alter the treatment experience? Inmates' perceptions of telemental health versus face-to-face treatment modalities. *Journal of Consulting and Clinical Psychology, 76*(1), 158–162. https://doi.org/10.1037/0022-006X.76.1.158.

Morland, L. A., Greene, C. J., Rosen, C. S., Foy, D., Reilly, P., Shore, J., … Frueh, B. C. (2010). Telemedicine for anger management therapy in a rural population of combat veterans with posttraumatic stress disorder: A randomized noninferiority trial. *Journal of Clinical Psychology, 71*(7), 855–863. https://doi.org/10.4088/jcp.09m05604blu.

Morland, L. A., Mackintosh, M., Greene, C. J., Rosen, C. S., Chard, K. M., Resick, P., & Frueh, B. C. (2014). Cognitive processing therapy for posttraumatic stress disorder delivered to rural veterans via telemental health: A randomized noninferiority clinical trial. *Journal of Clinical Psychiatry, 75*(5), 470–476. https://doi.org/10.4088/jcp.13m08842.

Morland, L. A., Mackintosh, M., Rosen, C. S., Willis, E., Resick, P., Chard, K., & Frueh, B. C. (2015). Telemedicine versus in-person delivery of cognitive processing therapy for women with posttraumatic stress disorder: A randomized noninferiority trial. *Depression and Anxiety, 32*, 811–820. https://doi.org/10.1002/da.22397.

Munder, T., & Barth, J. (2018). Cochrane's risk of bias tool in the context of psychotherapy outcome research: Considerations of how to conduct meta-analyses in psychological interventions. *Psychotherapy Research, 28*(3), 347–355. https://doi.org/10.1080/10503307.2017.1411628.

Myers, K., & Turvey, C. (Eds.). (2012). *Telemental health: Clinical, technical, and administrative foundations for evidence-based practice*. Newnes.

Nelson, E., Barnard, M., & Cain, S. (2003). Treating childhood depression over videoconferencing. *Telemedicine and e-health., 9*, 49–55. https://doi.org/10.1089/153056203763317648.

Norcross, J. C., Pfund, R. A., & Prochaska, J. O. (2013). Psychotherapy in 2022: A Delphi poll on its future. *Professional Psychology: Research and Practice, 44*(5), 363. https://doi.org/10.1037/a0034633.

Nordal, K. C. (2015). Embracing telepsychology. *Monitor on Psychology, 46*, 11–64.

Ohannessian, R., Duong, T. A., & Odone, A. (2020). Global telemedicine implementation and integration within health systems to fight the COVID-19 pandemic: A call to action. *The Journal of Medical Internet Research - Public Health & Surveillance, 6*(2), e18810. https://doi.org/10.2196/18810.

O'Reilly, R., Bishop, J., Maddox, K., Hutchinsons, L., Fisman, M., & Takhar, J. (2007). Is telepsychiatry equivalent to face-to-face psychiatry? Results from a randomized controlled equivalence trial. *Psychiatric Services, 58*(6), 836–843. https://doi.org/10.1176/ps.2007.58.6.836.

Osenbach, J. E., O'Brien, K. M., Mishkind, M., & Smolenski, D. J. (2013). Synchronous telehealth technologies in psychotherapy for depression: A meta-analysis. *Depression and Anxiety, 30*(11), 1058–1067. https://doi.org/10.1002/da.22165.

Porcari, C. E., Amdur, R. L., Koch, E. I., Richard, D. C., Favorite, T., Martis, B., & Liberzon, I. (2009). Assessment of post-traumatic stress disorder in veterans by videoconferencing and by face-to-face methods. *Journal of Telemedicine and Telecare, 15*, 89–94. https://doi.org/10.1258/jtt/2008/080612.

R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. URL http://www.R-project.org/.

Raudenbush, S. W. (2009). Analyzing effect sizes: Random-effects models. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 295–316).

Raudenbush, S. W., & Bryk, A. S. (1985). Empirical Bayes meta-analysis. *Journal of Educational Statistics, 10*(2), 75–98. https://doi.org/10.3102/10769986010002075.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods (Vol. 1)*. Sage.

Rees, C. S., & Maclaine, E. (2015). A systematic review of videoconference-delivered psychological treatment for anxiety disorders. *Australian Psychologist, 50*(4), 259–264. https://doi.org/10.1111/ap.12122.

Ruskin, P. E., Silver-Aylaian, M., Kling, M. A., Reed, S. A., Bradham, D. D., Hebel, J. R., … Hauser, P. (2004). Treatment outcomes in depression: Comparison of remote treatment through telepsychiatry to in-person treatment. *American Journal of Psychiatry, 161*(8), 1471–1476.

Salmoiraghi, A., & Hussain, S. (2015). A systematic review of the use of telepsychiatry in acute settings. *Journal of Psychiatric Practice, 21*(5), 389–393. https://doi.org/10.1097/PRA.0000000000000103.

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomized trials. *British Medical Journal, 340* (7748), 698–702. https://doi.org/10.1136/bmj.c332.

Shadish, W. R., & Haddock, C. K. (2009). Combining estimates of effect size. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 257–277) (Russell Sage Foundation).

Shulman, M., John, M., & Kane, J. M. (2017). Home-based outpatient telepsychiatry to improve adherence with treatment appointments: A pilot study. *Psychiatric Services, 68*(7), 743–746. https://doi.org/10.1176/appi.ps.201600244.

Simpson, S. (2001). The provision of a telepsychology service to Shetland: Client and therapist satisfaction and the ability to develop a therapeutic alliance. *Journal of Telemedicine and Telecare, 7*(5), 34–36. https://doi.org/10.1177/1357633X010070S114.

Simpson, S., Bell, L., Knox, J., & Mitchell, D. (2005). Therapy via videoconferencing: A route to client empowerment? *Clinical Psychology and Psychotherapy, 12*, 156–165. https://doi.org/10.1002/cpp.436.

Simpson, S., Guerrini, L., & Rochford, S. (2015). Telepsychology in a university psychology clinic setting: A pilot project. *Australian Psychologist, 50*, 285–291. https://doi.org/10.1111/ap.12131.

Sloan, D. M., Gallagher, M. W., Feinstein, B. A., Lee, D. J., & Pruneau, G. M. (2011). Efficacy of telehealth treatments for posttraumatic stress-related symptoms: A meta-analysis. *Cognitive Behaviour Therapy, 40*(2), 111–125. https://doi.org/10.1080/16506073.2010.550058.

Sproch, L. E., & Anderson, K. P. (2019). Clinician-delivered teletherapy for eating disorders. *Psychiatric Clinics of North America, 42*(2), 243–252. https://doi.org/10.1016/j.psc.2019.01.008.

Stain, H. J., Payne, K., Theinel, R., Michie, P., Carr, V., & Kelly, B. (2011). The feasibility of videoconferencing for neuropsychological assessment rural youth experiencing early psychosis. *Journal of Telemedicine and Telecare, 17*, 328–331. https://doi.org/10.1258/jtt.2011.101015.

Stead, A., & Vinson, M. (2019). Cognitive assessment using face-to-face and videoconferencing methods. *Nursing Older People.*. https://doi.org/10.7748/nop.2019.e1160.

Stubbings, D. R. (2012). *The effectiveness of videoconference-based cognitive-behavioural therapy*. Perth, Australia: Curtin University.

Stubbings, D. R., Rees, C. S., Roberts, L. D., & Kane, R. T. (2013). Comparing in-person to videoconference-based cognitive behavioral therapy for mood and anxiety disorders: Randomized controlled trial. *Journal of Medical Internet Research, 15*(11), 169–184.

Sutton, A. J. (2009). In H. Cooper, L. Hedges, & J. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed.). New York: Russell Sage Foundation.

Temple, V., Drummond, C., Valiquette, S., & Jozsvai, E. (2010). A comparison of intellectual assessments over video conferencing and in-person for individuals with ID: Preliminary data. *Journal of Intellectual Disability Research, 54*(6), 573–577. https://doi.org/10.1111/j.1365-2788.2010.01282.x.

Tse, Y. J., Mccarty, C. A., Stoep, A. V., & Myers, K. M. (2015). Teletherapy delivery of caregiver behavior training for children with attention-deficit hyperactivity disorder. *Telemedicine and e-Health, 21*(6), 451–458. https://doi.org/10.1089/tmj.2014.0132.

Tuerk, P. W., Yoder, M., Ruggiero, K. J., Gros, D. F., & Acierno, R. (2010). A pilot study of prolonged exposure therapy for posttraumatic stress disorder delivered via telehealth technology. *Journal of Traumatic Stress, 23*(1), 116–123. https://doi.org/10.1002/jts.20494.

United Kingdom National Health Service. The NHS long term plan. (August 21, 2019). https://www.longtermplan.nhs.uk.

Vahia, I. V., Ng, B., Camacho, A., Cardenas, V., Cherner, M., Depp, C. A., … Agha, Z. (2015). Telepsychiatry for neurocognitive testing in older rural latino adults. *The American Journal of Geriatric Psychiatry, 23*(7), 666–670. https://doi.org/10.1016/j.jagp.2014.08.006.

Viechtbauer, W. (2010). Conducting meta-analyses in r with the metafor package. *Journal of Statistical Software, 36*(3), 1–48.

Viechtbauer, W., & Cheung, M. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods, 1*(2), 112–125. https://doi.org/10.1002/jrsm.11.

Wadsworth, H. E., Galusha-Glasscock, J. M., Womack, K. B., Quiceno, M., Weiner, M. F., Hynan, L. S., … Cullum, C. M. (2016). Remote neuropsychological assessment in rural american indians with and without cognitive impairment. *Archives of Clinical Neuropsychology, 31*(5), 420–425. https://doi.org/10.1093/arclin/acw030.

Wierwille, J. L., Pukay-Martin, N. D., Chard, K. M., & Klump, M. C. (2016). Effectiveness of PTSD telehealth treatment in a VA clinical sample. *Psychological Services, 13*(4), 373–379. https://doi.org/10.1037/ser0000106.

Wong, L., Martin-Khan, M., Rowland, J., Varghese, P., & Gray, L. C. (2012). The Rowland university dementia assessment scale (RUDAS) as a reliable screening tool for dementia when administered via videoconferencing in elderly post-acute hospital patients. *Journal of Telemedicine and Telecare, 18*, 176–179. https://doi.org/10.1258/jtt.2012.2SFT113.

Wortmann, J. H., Jordan, A. H., Weathers, F. W., Resick, P. A., Dondaville, K. A., Hall-Clark, B., … Litz, B. T. (2016). Psychometric analysis of the PTSD Checklist-5 (PCL-5) among treatment-seeking military service members. *Psychological Assessment, 28* (11), 1392–1403. https://doi.org/10.1037/pas0000260.

Yoon, S., Lee, Y. J., Oh, S., Denninger, J. W., Kim, B. N., & Kim, J. H. (2018). A mobile videoconference-based intervention on stress reduction and resilience enhancement in employees: Randomized controlled trial. *Journal of Medical Internet Research, 20* (10), 1–19.

Yoshino, A., Shigemura, J., Kobayashi, Y., Nomura, S., Shishikura, K., Den, R., … Ashida, H. (2001). Telepsychiatry: Assessment of televideo psychiatric interview relaibilty with present- and next-generation internet infrastructures. *Acta Psychiatrica Scandinavica, 104*, 223–226. https://doi.org/10.1034/j.1600-0447.2001.00236.x.

Young, L. B. (2012). Telemedicine interventions for substance-use disorder: A literature review. *Journal of Telemedicine and Telecare, 18*(1), 47–53. https://doi.org/10.1258/jtt.2011.110608.

Yuen, E. K., Gros, D. F., Price, M., Zeigler, S., Tuerk, P. W., Foa, E. B., & Acierno, R. (2015). Randomized controlled trial of home-based telehealth versus in-person prolonged exposure for combat-related PTSD in veterans: Preliminary results. *Journal of Clinical Psychology, 71*(6), 500–512. https://doi.org/10.1002/jclp.22168.

Zarate, C. A., Weinstock, L., Cukor, P., Morabito, C., Leahy, L., Burns, C., & Baer, L. (1997). Applicability of telemedicine for assessing patients with schizophrenia: Acceptance and reliability. *Journal of Clinical Psychiatry, 58*(1), 22–25. https://doi.org/10.4088/JCP.v58n0104.

Zhou, X., Snoswell, C. L., Harding, L. E., Bambling, M., Edirippulige, S., Bai, X., & Smith, A. C. (2020). The role of telehealth in reducing the mental Health burden from COVID-19. *Telemedicine journal and e-health: the official journal of the American Telemedicine Association, 26*(4), 377–379. https://doi.org/10.1089/tmj.2020.0068.

Ashley B. Batastini received her PhD in counseling psychology at Texas Tech University. She is currently assistant professor in the Department of Counseling, Educational Psychology, and Research at the University of Memphis. Dr. Batastini was affiliated with the University of Southern Mississippi when this work was completed. Her professional interests include development of and research on mental health interventions for criminal justice-involved populations and best practices in forensic mental health assessment.

Peter Paprzycki received his Ph.D. in Research, Evaluation, Statistics and Measurement at the University of Toledo. He is a part-time Professor at the University of Toledo, and a statistical consultant at Mississippi Center for Clinical and Translational Research (MCCTR) and Grant Fundamentals, LCC. Dr. Paprzycki is a member of the Networking Urban Resources with Teachers and University to enRich Early Childhood Science

(NURTURES), a National Science Foundation funded Mathematics and Science Partnership (MSP) program.

Ashley C. T. Jones received her master's degree at Arizona State University. She is currently a doctoral student at the University of Southern Mississippi. Her professional interests include innovative technologies in correctional treatment, program evaluation, and research on jury decision-making.

Nina MacLean received her PhD in counseling psychology at Texas Tech University. She is currently a consulting forensic examiner at the Center for Forensic Psychiatry. Her professional interests include professional issues and best practices in forensic mental health evaluations.