Internet Interventions 25 (2021) 100399

Contents lists available at ScienceDirect

Internet Interventions

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

The Supportive Accountability Inventory: Psychometric properties of a measure of supportive accountability in coached digital interventions

Jonah Meyerhoff *, Shefali Haldar, David C. Mohr

Center for Behavioral Intervention Technologies (CBITs), Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, 750 North Lake Shore Drive, 10th Floor, Chicago, IL 60611, United States of America

ARTICLE INFO

Keywords:
Digital mental health
Behavioral intervention technology
Human support
mHealth

ABSTRACT

Background: One of the most widely used coaching models is Supportive Accountability (SA) which aims to provide intervention users with clear expectations for intervention use, regular monitoring, and a sense that coaches are trustworthy, benevolent, and have domain expertise. However, few measures exist to study the role of the SA model on coached digital interventions. We developed the Supportive Accountability Inventory (SAI) and evaluated the underlying factor structure and psychometric properties of this brief self-report measure.
Method: Using data from a two-arm randomized trial of a remote intervention for major depressive disorder (telephone CBT [tCBT] or a stepped care model of web-based CBT [iCBT] and tCBT), we conducted an Exploratory Factor Analysis on the SAI item pool and explored the final SAI's relationship to iCBT engagement as well as to depression outcomes. Participants in our analyses (n = 52) included those randomized to a receive iCBT, but were not stepped up to tCBT due to insufficient response to iCBT, had not remitted prior to the 10-week assessment point, and completed the pool of 8 potential SAI items.
Results: The best fitting EFA model included only 6 items from the original pool of 8 and contained two factors: Monitoring and Expectation. Final model fit was mixed, but acceptable ($\chi^2(4) = 5.24$, $p = 0.26$; RMSR = 0.03; RMSEA = 0.091; TLI = 0.967). Internal consistency was acceptable at $\alpha = 0.68$. The SAI demonstrated good convergent and divergent validity. The SAI at the 10-week/mid-treatment mark was significantly associated with the number of days of iCBT use ($r = 0.29$, $p = .037$), but, contrary to expectations, was not predictive of either PHQ-9 scores ($F(2,46) = 0.14$, $p = .89$) or QIDS-C scores ($F(2,46) = 0.84$, $p = .44$) at post-treatment.
Conclusion: The SAI is a brief measure of the SA framework constructs. Continued development to improve the SAI and expand the constructs it assesses is necessary, but the SAI represents the first step towards a measure of a coaching protocol that can support both coached digital mental health intervention adherence and improved outcomes.

1. Introduction

Digital mental health interventions (DMHIs) address behavioral health targets to support physical, behavioral, and mental health (Hollis et al., 2017). DMHIs have been implemented as web-based tools, teleconferencing, social media, and mHealth apps, (Aguilera, 2015) and have the potential to help individuals overcome barriers to accessing and engaging in mental health treatment (Borghouts et al., 2021; Schueller and Torous, 2020; Sorkin et al., 2021). Although DMHIs have demonstrably improved depression outcomes (Firth et al., 2017), consistent use and measuring adherence to DMHIs remains a challenge in real-world contexts (Donkin et al., 2011; Sieverink et al., 2017). While the overarching purpose of coaching is to improve outcomes, one of the primary intermediate goals of coaching is to increase adherence and engagement (Graham et al., 2019) — a similar metric to adherence that relies more heavily on objective markers of DMHI use rather than *intended* use. Coaching improves engagement by circumventing (1) usability problems (e.g., poor fit to user needs and preferences or poor functionality), (2) user motivation, and (3) skill implementation, enabling users to incorporate the DMHI into their everyday life despite design gaps (Schueller et al., 2017).

To overcome these challenges, coached DMHIs have been introduced

* Corresponding author at: Center for Behavioral Intervention Technologies, Northwestern University Feinberg School of Medicine, 750 North Lake Shore Drive, 10th Floor, Chicago, IL 60611, United States of America.
E-mail address: jonah.meyerhoff@northwestern.edu (J. Meyerhoff).

https://doi.org/10.1016/j.invent.2021.100399
Received 8 November 2020; Received in revised form 28 April 2021; Accepted 29 April 2021
Available online 4 May 2021

to integrate human support alongside the technology intervention (Schueller et al., 2017). A human coach is typically responsible for reminding patients about completing intervention components and helping them with technical or treatment troubleshooting. In mental health contexts, coached interventions involve regular weekly or monthly check-ins that take place via phone or email (Agboola et al., 2015; Stiles-Shields et al., 2019). For web- or mobile app-based interventions, coaches might also have access to data about the patient's use of the technology intervention to tailor these phone check-ins.

Many studies have successfully demonstrated the ability of coached DMHIs to increase engagement, reduce clinical trial dropout rates (Torous et al., 2020), and improve outcomes for depression and anxiety (Andersson and Cuijpers, 2009; Linardon et al., 2019; Mohr et al., 2019b). However, the nature and implementation of coached DMHIs can vary greatly (Hermes et al., 2019), highlighting the importance of measuring and evaluating coached DMHIs. Supportive Accountability (SA) is a widely used model for coaching in digital health (Chhabria et al., 2020; Dennison et al., 2014; Pilutti et al., 2014) that provides a theoretical basis for increasing the consistency, reach, and effectiveness of coached DMHIs. Its success depends on the patient's perception of characteristics–such as trustworthiness, benevolence, and expertise–of the human who serves as the coach (Mohr et al., 2011).

Previous work has examined similar psychological processes shown to be important factors with the potential to impact psychotherapy outcomes such as how credible a treatment is and how well one expects it to work (Borkovec and Nau, 1972; Newman and Fisher, 2010; Smeets et al., 2008), the strength of a patients' motivation to engage in treatment (Pelletier et al., 1997; Vogel et al., 2006), and how aligned the therapist and patient are (Bordin, 1979; Flückiger et al., 2018; Horvath and Greenberg, 1989). Perhaps the most well studied of these psychotherapy processes is the therapeutic alliance which was first developed as a model of the relational dynamic that develops between a therapist and a patient (Bordin, 1979). It relies on the patient and therapist mutually agreeing on therapeutic goals, fulfillment of therapeutic tasks (e.g., thought records), and establishing a mutual trust. While therapeutic alliance and SA are similar in that they are both models of human connection that are important for maintaining adherence in, and efficacy of, psychotherapy, and they both involve patient perceptions of how closely a provider matches their needs, SA differs in a number of important ways. First the SA model is specifically designed to guide digital coaches, whose role is to support engagement and adherence to a digital intervention, rather than deliver the active ingredients of an intervention. Second, the SA model is built on the principles that a digital coach is perceived as legitimate, forms a bond with users, and facilitates accountability, whereas therapeutic alliance may achieve some of these goals, it is heavily focused on "goodness of fit" between patient and therapist (Bordin, 1979). SA, in contrast, relies on the perception that a coach is legitimate (has expertise, is trustworthy, benevolent, and maintains a reciprocal relationship), establishes a clear bond with the user, and establishes structures for accountability (e.g., sets clear measurable expectations, goals, monitors use and performance, is process oriented, etc.). While these two constructs have similarities, each is unique and aimed at affecting different aspects of the patient or user relationship to a provider or coach. Therapeutic alliance has been sparsely studied in DMHIs, but the few studies that have measured therapeutic alliance in the context of digital interventions have found that it can be effectively established via DMHIs (Flückiger et al., 2018; Pihlaja et al., 2018; Sucala et al., 2012; Wehmann et al., 2020). Expanding from this literature, recent work has proposed the construct of a digital therapeutic alliance between users of a digital therapeutic and the technology itself (Henson et al., 2019). Importantly, we are only able to understand the role of therapeutic alliance and other psychotherapy processes such as treatment credibility and expectancy, and clients' motivation for treatment as they relate to psychotherapy and DMHIs because well-established measures of working alliance have been developed and are used across settings today (Devilly and Borkovec, 2000; e.g., Hatcher and Gillaspy, 2006; Horvath and Greenberg, 1989; Pelletier et al., 1997). To adequately study the role of the SA model on coached digital interventions, a measure is needed to assess SA constructs as they relate to coached DMHIs.

We address the gap in this literature by presenting the Supportive Accountability Inventory (SAI), a means to help evaluate coached interventions, increase their use, and improve their consistent implementation. In this paper, we evaluate the underlying factor structure and psychometric properties of a brief self-report measure of perceived support and accountability provided through digital intervention coaching. We also evaluate the relationship between perceived support and accountability and intervention engagement. Our findings can help researchers with increasing the efficacy of coached DMHIs and ultimately lead to improved mental health outcomes.

2. Material and methods

2.1. Development of supportive accountability inventory (SAI) items

Supportive Accountability Inventory items were initially developed based on extensive literature review and clinical experience. All items were generated by four psychologists with expertise in coached digital mental health intervention development and research. Items based on the SA model of coaching (Mohr et al., 2011) were generated and refined via consensus discussion. Items were piloted in a field trial of coached web-based cognitive behavioral therapy for depression (Schueller and Mohr, 2015). Participants in the field trial responded to a web-based questionnaire that included 8 candidate SAI items (Table 1). Participants were asked to mentally fill in the blanks in each item with their coach's name. Items were rated on a 1–7 Likert scale ranging from Strongly Disagree (1) to Strongly Agree (7). Total scores were calculated by summing items, higher total scores indicated higher perceived supportive accountability.

2.2. Participants and procedures

This study is a secondary analysis of data from a two-arm randomized trial of a remote intervention for major depressive disorder (MDD; the methodological and procedural details are reported in Mohr et al., 2019a). Briefly, participants were recruited between the winter of 2015 and spring of 2017 from public-facing advertisements and through institutional medical networks. To participate, individuals were required to be at least 18 years old, have a current major depressive episode based on the Mini International Neuropsychiatric Interview (MINI; Sheehan et al., 1998), have a score of at least 12 on the Quick inventory of Depressive Symptomology – Clinician rated (QIDS-C; Rush et al., 2006), and have access to a web-enabled device, be able to read, speak, and understand English. Participants were excluded if they were experiencing a psychiatric or medical condition for which participation was contraindicated (e.g., suicidal ideation), and had recently (within the last 2 weeks) initiated psychotherapy or pharmacotherapy for their

Table 1
SAI item pool.

Item number	Item content
1	_______ will notice if I go above and beyond to find creative ways to use [platform name].
2	_______ does not really keep track of progress I make in using [platform name].
3	If I use [platform name] less frequently than is expected, I will feel the need to justify my reasons why to _____.
4	_______ is aware of and notices when I use [platform name].
5	_______ is aware of how I have completed the [platform name] tools.
6	_______ expects that I will be very consistent in using [platform name].
7	_______ expects I follow-up on the goals set during our weekly calls.
8	_______ will notice if I get better at using the [platform name] tools.

depressive symptoms (full inclusion and exclusion criteria is described in detail in Mohr et al., 2019a). All study procedures were approved by the Northwestern University Institutional Review Board, and all participants provided their informed consent to participate prior to the initiation of any study procedures.

Participants were randomized to receive either (1) telephone cognitive behavioral therapy (tCBT) or (2) stepped care for depression until full remission was reached [patient health questionnaire – 9 (PHQ-9; Kroenke et al., 2001) < 5 for two weeks] or until participants received 20 weeks of treatment. In stepped care, individuals received web-based CBT (iCBT) with coaching (Tomasino et al., 2017) and were "stepped up" to tCBT if depressive symptoms were predictive of an insufficient response to iCBT [PHQ-9 $\geq$ 17 from weeks 4–8, PHQ-9 $\geq$ 13 from weeks 9–13, or PHQ-9 $\geq$ 9 after week 13]. For additional procedural details, see Mohr et al. (2019a). Coaching was defined by a manualized protocol (Tomasino et al., 2017), and consisted of an initial engagement call, usually lasting approximately 30–40 min. These calls were followed by brief 10–15 min contacts that supported use of the iCBT platform as well as several (2–3) digital message contacts per week. After three weeks of telephone-based coaching contacts, participants were able to opt for message-only coaching and telephone coaching as-needed. Coaches received training and weekly supervision, which included review of audiotaped calls and message logs.

During the treatment phase, participants completed assessments at baseline, mid-treatment (week 10), and end of treatment (EOT) or week 20 (whichever came first). Participants who achieved full remission or were stepped up to tCBT prior to week 10, did not complete the mid-treatment/week 10 assessments.

For the present study, we focused our analyses on participants who were administered the pool of 8 potential SAI items (n = 52). The 8 potential SAI items were administered at the 10-week (mid-treatment) timepoint and therefore was only administered to individuals randomized to the stepped care treatment arm, received iCBT, but had not stepped up to tCBT, and had not remitted prior to the 10-week assessment point. This allowed us to examine the factor structure of the SAI in a sample who received regular coaching on the iCBT platform and reported on their experiences of coaching at the 10-week timepoint.

2.3. Measures

Demographics – Participants were administered a brief survey assessing demographic information at baseline, age and gender were taken into consideration for the purposes of these secondary analyses. Demographic questionnaires were administered only at baseline.

The *Patient Health Questionnaire – 9* (PHQ-9; Kroenke et al., 2001) is a 9-item measure assessing DSM-5 (American Psychiatric Association and American Psychiatric Association (Eds.), 2013) criteria for depression along 9 possible symptoms including mood, anhedonia, appetite changes, and suicidal ideation. A supplementary 10th item assesses the degree of impairment an individual experiences as a result of their depressive symptoms. The measure assesses symptom severity over the past two weeks. The PHQ-9 was administered at each assessment timepoint (baseline, mid-treatment/10-weeks, EOT/20-weeks).

The *Quick Inventory of Depressive Symptomology – Clinician version* (QIDS-C; Rush et al., 2006) is a clinician-rated depression inventory. The QIDS-C was administered at baseline, mid-treatment/10-weeks, EOT/20-weeks.

The *Working Alliance Inventory-Short form Revised* (WAI-SR; Hatcher and Gillaspy, 2006) is a brief 12-item self-report measure of alignment and working alliance. The WAI-SR was administered at mid-treatment/10-weeks and EOT/20-weeks.

The *Client Motivation for Therapy Scale* (CMTS; Pelletier et al., 1997) is a 24-item self-report assessment of the intrinsic motivation for therapy and was administered at mid-treatment/10-weeks.

The *Credibility and Expectancy Questionnaire* (CEQ; Devilly and Borkovec, 2000) is a 6-item questionnaire that assesses what patients *think* about the effectiveness of a treatment and how effective they *feel* that it will be. The CEQ was administered at each timepoint (baseline, mid-treatment/10-weeks, EOT/20-weeks).

The *Supportive Accountability Inventory* (SAI), as previously noted, was explored using 8 core items assessing the clients' perceptions of Supportive Accountability from their digital intervention coach. Most items asked participants to mentally fill-in-the-blank with then name of their coach. Items were rated on a Likert-scale ranging from 1, 'strongly disagree,' to 7, 'strongly agree.' The pool of 8-items was administered to participants at mid-treatment/10-weeks and EOT/20-weeks.

2.4. Statistical analyses

Our primary aims for this study were to explore the underlying factor structure of the SAI and also determine the extent to which the SAI was associated with intervention use across treatment. To achieve these two aims, we first explored the number of factors to include in an exploratory factor analysis (EFA) that used all SAI items at mid-treatment/10-weeks to refine the pool of SA items and reveal the underlying factor structure of the SAI. While most EFA's require splitting the sample into two, in order to cross validate findings from the EFA in a confirmatory factor analysis (CFA), we were unable to conduct this CFA step due to the small sample size. This is further addressed in the limitations section. Prior to running an EFA, we conducted several diagnostic tests to determine the appropriateness of conducting a factor analysis. We ran Bartlett's test of sphericity, which tests whether the correlation matrix in question is significantly different from an identity matrix. This revealed that the original item correlation matrix was significantly different from an identity matrix (p < .0001). We then conducted a Kaiser Meyer Olkin (KMO) test to tell us if sampling was sufficient. KMO test statistic was >0.6 suggesting that the sampling was mediocre, but that analyses could proceed, but results should be interpreted with caution as there may be substantial partial correlations that could impact a factor analysis (Cerny and Kaiser, 1977). Finally we ran a parallel analysis (Horn, 1965) which revealed 2 as the optimal number of factors. Because data were not normally distributed, we used a principle axis factor extraction method in the EFA. Additionally, because latent factors were expected to correlate with one another, an oblique rotation method was used (promax). We examined standardized factor loadings to remove items with significant cross-loadings (0.3 or higher) and/or low individual loadings (<0.3).

Additionally, we examined the SAI's convergent and divergent validity with measures of related constructs to SA. These related constructs included working alliance (WAI-SR), treatment credibility and expectancy (CEQ), and motivation for treatment (CMTS). We expected that the SAI would demonstrate convergent validity with regard to working alliance, but divergent validity with regard to treatment credibility and motivation for treatment.

To examine the associative relationship SAI scores had with platform use across the digital intervention, we examined the correlation (Pearson) between SAI scores at mid-treatment/week 10 of the coached digital intervention and the number of days patients used the intervention.

We also conducted a series of linear regressions (using both the PHQ-9 and QIDS-C) to determine if SAI scores at mid-treatment/Week 10, were predictive of depression scores at EOT/Week 20, adjusting for baseline depression.

All data analyses were formed using R version 3.6.1 (R Core Team, 2019) with the following packages: tidyverse (Wickham and RStudio, 2019), dplyr (Wickham and Francois, 2015), reshape2 (Wickham, 2007), corrplot (Wei et al., 2017), Hmisc (Harrell Jr., 2020), psych (Revelle, 2020), GPArotation_2014 (Bernaards and Jennrich, 2014), gridExtra (Auguie and Antonov, 2017), and ggplot2 (Wickham, 2009).

3. Results

3.1. Participant characteristics

52 participants were included in these analyses as they completed all items of the SAI at mid-treatment/the 10-week point and were not stepped up to tCBT. Participants were a mean age of 37.94 (SD = 13.43). Most participants identified as women (75%; 39/52), the remaining participants identified as men (25%; 13/52). The sample makeup was 82.6% (43/52) White, 3.8% (2/52) Black, 3.8% (2/52) Asian, and 9.6% (5/52) identifying as more than one race. A minority of the sample identified as Hispanic or Latinx 3.8% (2/52).

3.2. Missing data

Missing data were minimal in this sample; one individual did not complete all SAI items at mid-treatment and was excluded from analyses. Three participants who did have SAI scores at mid-treatment/10 weeks did not report PHQ-9 scores at post-treatment, and two did not report QIDS-C scores at post-treatment. Given the low rate of missing data in this sample, missingness was not significantly associated with SAI score, baseline depression score, or participant demographics.

3.3. Exploratory factor analysis

SAI scores ($n = 52$) were slightly non-normally distributed (skewness = −0.52; kurtosis = −0.15). Subsequently, we used a principal factor extraction method with promax rotation. Factor analysis revealed the best fitting model included only 6 items from the original pool of 8. These 6 items loaded onto two factors (Fig. 1): Monitoring and Expectation. Final model fit was mixed, but acceptable ($\chi^2(4) = 5.24$, p=0.26; RMSR =0.03; RMSEA = 0.091; TLI = 0.967). The internal consistency of the overall measure was acceptable at α = 0.68, as was the monitoring (α = 0.66) and expectation subscales (α = 0.70). The monitoring subscale comprises items assessing whether the user perceived that the coach is paying attention to a user's individual use or, and progress through, the digital intervention. The expectations subscale is made up of items assessing a user's sense that the digital coach maintains specific expectations around the individual user's engagement with the intervention platform and around a user's concrete actions taken towards treatment-related goals.

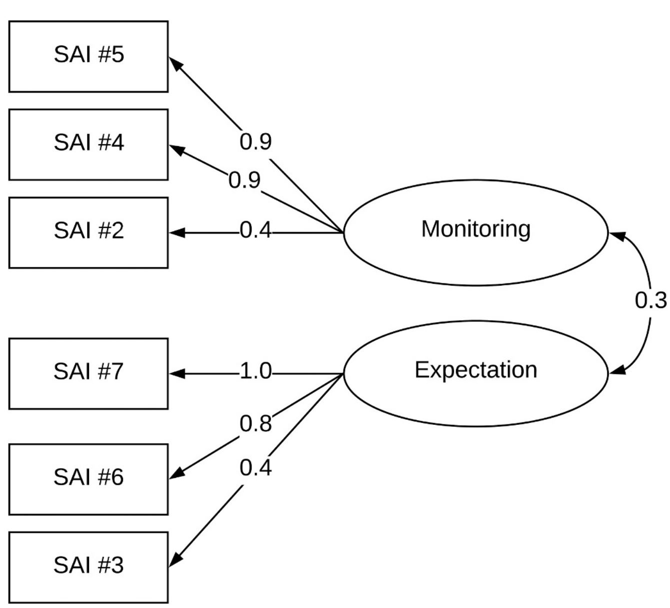

Fig. 1. EFA Model Results.
SAI item numbers correspond to those listed in Table 1.

3.4. Convergent and divergent validity

The SAI was moderately correlated ($n = 50$, $r = 0.35$, $p = .011$) with the WAI-SR and the CEQ ($n = 52$, $r = 0.34$, $p = .012$), but was not significantly correlated ($n = 52$, $r = 0.22$, $p = .11$) with the CMTS scale, demonstrating divergent validity.

3.5. Supportive accountability and platform use

The SAI total score at mid-treatment/week-10 was significantly associated with the number of days logged into the digital mental health intervention ($r = 0.29$, $p = .037$). Only the Expectation subscale was also associated with number of days the platform was used across the intervention period ($r = 0.30$, $p = .033$). The Monitoring subscale was not associated with the number of days logged in ($r = 0.13$, $p = .34$). Results suggest that SAI – in particular the Expectations subscale – accounted for a medium amount of the total variance in number of days logged in.

3.6. Supportive accountability and depression outcomes

Contrary to expectations, SAI scores at mid-treatment were not predictive of either PHQ-9 scores ($F(2,46) = 0.14$, $p = .89$) or QIDS-C scores ($F(2,46) = 0.84$, $p = .44$) at post-treatment, adjusted for baseline scores, respectively.

4. Discussion

This study examined the underlying factor structure of a brief measure of Supportive Accountability, the SAI, as well as the measure's relationship to important relevant constructs such as engagement and symptom reduction. We found that the SAI was best fit using 6-items from an original pool of 8, the items that were dropped loaded poorly onto the two-factor EFA model. The SAI comprises a monitoring factor assessing whether a user believes that a digital coach pays attention to their use of, and progress through, the digital intervention, and an expectations factor which measures a user's thoughts that a digital coach maintains expectations of a user's engagement with the digital intervention and effort towards treatment-related goals. Model fit for the SAI was adequate, and internal consistency was acceptable. We found the SAI attained good convergent validity with constructs like working alliance as well as treatment credibility and expectancy, but was also divergent enough to add meaningful value. Importantly, we found that the SAI was not associated with psychotherapy processes hypothesized to be poorly related to the measure such as patient motivation for treatment. These findings suggest that the SAI provides a good starting point to begin measuring the Supportive Accountability model as delivered in coached digital interventions.

We found that the SAI overall, driven by the expectations subscale, was associated with patient engagement across the trial period in the form of the number of days logged in. This finding highlights the importance of coaches setting clear expectations for the user to engage with the intervention platform. Broadly in psychotherapy, two expectation types exist. The first is expectations about the outcome of therapy, or how much a user expects their symptoms will be reduced through psychotherapy (Constantino et al., 2011). The second is expectations about the therapeutic process, in other words, what the coach expects patients to actually do during the therapeutic process (Lerner and Tetlock, 1999; Mohr et al., 2011). The Supportive Accountability model relies primarily on the early establishment of process expectations to ensure coaches hold users accountable throughout the intervention process, not just after early non-engagement in the intervention. If accountability happens only after a period of user non-engagement, cognitive dissonance theory suggests that a person's defensive position of their non-engagement will be strengthened, which may interfere with the therapeutic change process (Mohr et al., 2011). The SAI's expectations subscale primarily focuses on these process expectations and plays an important role in intervention engagement.

Surprisingly, the monitoring subscale was – in our study – unrelated to the number of days logged in. While this may be the result of our small sample size, it may also suggest that within the context of this particular digital depression intervention, the participants' perception of intervention usage monitoring by the coach was a less potent driver of behavior relative to process expectations. This could stem from the delicate nature of performance monitoring in a depression intervention. While monitoring is an important part of the Supportive Accountability model, heavy-handed monitoring can be seen by users and patients as controlling and have the opposite effect: decreased engagement (Mohr et al., 2011). Therefore, while it is possible that monitoring effects on engagement were less potent than the expectation effects in our study, a more likely explanation is that explicit monitoring behaviors by coaches were too low to yield engagement effects due to an overly conservative approach in an effort to prevent monitoring from becoming perceived as overbearing. Thus, while performance monitoring is an important construct, coaches may need to deliver this Supportive Accountability component more liberally. Alternatively, it is possible that the 3 monitoring items on the SAI were insufficiently specific to capture the true effects of monitoring alone on patient engagement.

Another important finding was that we were unable to detect an association between the SAI and our primary symptom outcome, depression severity. This finding affirms those of a recently published Supportive Accountability Measure (SAM; Chhabria et al., 2020), which was specific to weight loss interventions. The SAM was associated with improved adherence to the weight loss interventions, but was not associated with weight change – the study's primary outcome. This suggests that elements of measures of Supportive Accountability (i.e., the SAM, and the SAI's expectation subscale) are associated with changes in intervention engagement or usage, but not necessarily with improvement in the clinical outcomes.

One reason that supportive accountability may be associated with intervention usage but not outcomes is that the relationship between DMHI use and clinical outcome is weak at best, with many studies finding no significant relationship at all (Donkin et al., 2013; Fuhr et al., 2018; Mohr et al., 2010). When we posit a dose-response, we assume a unidirectional relationship, in which more use leads to better outcomes. This assumption is likely wrong for at least two reasons. First, the relationship between intervention use and outcomes is likely complex and bidirectional. While some exposure to a DMHI can reduce symptom severity, level of symptom severity may also affect usage in the opposite direction. That is, people may use the intervention more when feeling worse and decrease usage as when they feel better (Mohr et al., 2010). These complex, bidirectional relationships, which likely occur over days and weeks, may obscure the relationship between use and symptom change in coarser analyses comparing use and outcome over an entire treatment period.

Another reason for the weak relationship between use data and clinical outcome is that these analyses assume that the number of times a person utilizes an intervention is a measure of engagement in the treatment. However, it is increasingly clear that engagement is not just a question of how much a person uses the DMH tool, but also *how* the person uses it. Emerging conceptualizations of engagement with digital interventions include not only the quantity of use, but also how involved people are cognitively, affectively, and behaviorally with the intervention (Donkin and Glozier, 2012; Kelders et al., 2020b). Someone who uses an application in a perfunctory manner would be less likely to benefit than someone who thinks about what they are learning, is emotionally engaged, and acts on what they are learning.

The growing body of research on coaching models focused on maintaining adherence, such as Supportive Accountability, suggests that they are effective at improving adherence, but not necessarily at improving outcomes. Coached interventions have generally reported better outcomes, adherence rates, and retention than self-guided digital mental health interventions (Baumeister et al., 2014; Baumel et al., 2019; Torous et al., 2020). One recent report (Josephine et al., 2017) notes the absence of a dose-response to coaching contact with users, suggesting that more coaching contact does not result in ever improving outcomes. Therefore, while there may be evidence that the presence of coaching provides some value, associations between coaching and outcomes within any given intervention and among studies of narrow populations are currently, at best, low. We view this issue as less about whether or not coaching ought to be provided as part of a DMHI, but more about understanding the best focus of that coaching. It is conceivable that adherence is not the right focus for coaching, but a more effective focus could be coaching that targets aspects of engagement (e.g., cognitive and affective, aspects of engagement; Kelders et al., 2020a) and intervention mechanisms of action. While supporting DMHI use is an important objective for coaches, expanding the coaching protocols that include goals related to how a person uses the tools may improve outcomes. The Efficiency Model of coaching support (Schueller et al., 2017) posits five potential failure points in DMHIs which a coach can address, including helping the patient manage usability problems with the tools, overcome motivational or attitudinal barriers to use, understand how to use the tools most effectively, find tools that match the patient's needs and preferences, and implement new skills into the context of the patient's life. Thus, supportive accountability likely represents a necessary component of coaching, but may not be sufficient by itself to promote symptom improvement. It is critical to measure not just what the coach is doing, but what the user is receiving from coaching, therefore, measurement tools, such as the SAI, could be expanded to measure a broader range of coaching objectives, which would allow us to develop a better understanding of the components of coaching that are effective at improving engagement and clinical outcomes.

4.1. Limitations

This study has a number of limitations that should be considered in interpreting these findings. First and foremost, while the sample size used in this study is likely robust in the face of common distortions (de Winter et al., 2009), it is nonetheless small. Future studies should verify results of the present study by conducting both an EFA and CFA in larger samples. Another limitation to the current study is that the EFA and subsequent exploratory analyses were conducted using the same sample. It is desirable to use separate samples to conduct an EFA, CFA, and subsequent exploratory analyses of mechanisms and outcomes. Future studies that include larger sample sizes may consider conducting each of these data analytic steps in separate samples. The item pool for this study was relatively small, though developed with user feedback. It will be useful to increase the pool of items to ensure the pool has items that are intended to map onto each of the overarching constructs in the Supportive Accountability model (e.g., Bond, Accountability, Legitimacy) as well as the specific constructs that make up each model component (i.e., accountability comprises social presence, expectations, process focus, goal setting, and performance monitoring). Expanded item pools and improved existing SAI items as well as larger sample sizes may also result in improved model fit and internal consistency. Second, because these data came from a DMHI offered in a stepped care model, participants who responded poorly were stepped up to tCBT prior to the administration of the SAI, similarly, those who remitted prior to the administration of the SAI also did not complete the SAI. We, therefore, cannot draw conclusions about the perceived SA of individuals who remitted quickly (i.e., prior to the 10-week mark) or those who did not have an adequate treatment response and were stepped up to tCBT. Additionally, administering the SAI to only the sample included in this study may have eliminated those with the least adherence to the DMHI, limited variability in SAI responses, and restricted the range of PHQ-9 and QIDS-C outcome scores. Thus, it is possible that the SAI may have had some relationship to outcome among a broader range of participants, however, given the Chhabria et al.'s (2020) findings and the poor relationship between usage and outcome generally, we still feel our comments on the need for an expanded view of coaching, such as the Efficiency Model (Schueller et al., 2017), are valid. Finally, we emphasize that the SAI is a measure of perceived expectations and monitoring; increasing the breadth of items administered and broadening the sample are necessary next steps in the development of the SAI. Coaches in this study used a manualized protocol and were closely supervised, limiting the variability of coach behavior. The relationships between the SAI and usage or outcomes could be larger in more naturalistic settings where operationalization of coaching is less controlled.

5. Conclusion

The SAI is a brief and adequately performing measure of the Supportive Accountability model constructs. The underlying two factor structure assesses the effect of a coach's performance monitoring and expectation setting. The SAI is associated with intervention use in the context of a digital mental health intervention for depression. Effects of the SAI on outcomes were not significant, as a result, continued development work to improve the SAI and expand coaching protocols is required to further understand the relationship between DMHI engagement and clinical outcomes. The SAI represents the first step towards a measure of the Supportive Accountability model in coached digital interventions.

Funding

This work was supported by a grant from the National Institute of Mental Health (R01MH095753-03) to David C. Mohr. David C. Mohr had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Jonah Meyerhoff and Shefali Haldar are supported by a grant from the National Institute of Mental Health (T32MH115882). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The NIMH had no part in study design, hypothesis generation, data collection, processing, or analysis, or interpretation of results. The NIMH had no role in authoring this publication.

Data availability statement

The source code for data preparation and analysis is available upon request to the corresponding author JM. The de-identified self-report data may be made available upon request to DCM, with required data use agreements.

Declaration of competing interest

David C. Mohr has accepted consulting fees from Apple Inc., Pear Therapeutics, Otsuka Pharmaceuticals, and the One Mind Foundation. He also has an ownership interest in Adaptive Health, Inc. None of the other authors have competing interests to declare.

References

Agboola, S.O., Ju, W., Elfiky, A., Kvedar, J.C., Jethwani, K., 2015. The effect of technology-based interventions on pain, depression, and quality of life in patients with cancer: a systematic review of randomized controlled trials. J. Med. Internet Res. 17 (3), e65 https://doi.org/10.2196/jmir.4009.

Aguilera, A., 2015. Digital technology and mental health interventions: opportunities and challenges. Arbor 191 (771), a210. https://doi.org/10.3989/arbor.2015.771n1012.

American Psychiatric Association, American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental Disorders: DSM-5, 5th ed. American Psychiatric Association.

Andersson, G., Cuijpers, P., 2009. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn. Behav. Ther. 38 (4), 196–205. https://doi.org/10.1080/16506070903318960.

Auguie, B., & Antonov, A. (2017). gridExtra: Miscellaneous Functions for “Grid” Graphics (2.3) [Computer software]. https://CRAN.R-project.org/package=gridExtra.

Baumeister, H., Reichler, L., Munzinger, M., Lin, J., 2014. The impact of guidance on internet-based mental health interventions—a systematic review. Internet Interv. 1 (4), 205–215. https://doi.org/10.1016/j.invent.2014.08.003.

Baumel, A., Edan, S., Kane, J.M., 2019. Is there a trial bias impacting user engagement with unguided e-mental health interventions? A systematic comparison of published reports and real-world usage of the same programs. Transl. Behav. Med. 9 (6), 1020–1033. https://doi.org/10.1093/tbm/ibz147.

Bernaards, C., Jennrich, R., 2014. GPArotation: GPA Factor Rotation (2014.11-1) [Computer software]. https://CRAN.R-project.org/package=GPArotation.

Bordin, E.S., 1979. The generalizability of the psychoanalytic concept of the working alliance. Psychol. Psychother. Theory Res. Pract. 16 (3), 252–260. https://doi.org/10.1037/h0085885.

Borghouts, J., Eikey, E., Mark, G., Leon, C.D., Schueller, S.M., Schneider, M., Stadnick, N., Zheng, K., Mukamel, D., Sorkin, D.H., 2021. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J. Med. Internet Res. 23 (3), e24387 https://doi.org/10.2196/24387.

Borkovec, T.D., Nau, S.D., 1972. Credibility of analogue therapy rationales. J. Behav. Ther. Exp. Psychiatry 3 (4), 257–260. https://doi.org/10.1016/0005-7916(72)90045-6.

Cerny, B.A., Kaiser, H.F., 1977. A study of A measure of sampling adequacy for factor-analytic correlation matrices. Multivar. Behav. Res. 12 (1), 43–47. https://doi.org/10.1207/s15327906mbr1201_3.

Chhabria, K., Ross, K.M., Sacco, S.J., Leahey, T.M., 2020. The assessment of supportive accountability in adults seeking obesity treatment: psychometric validation study. J. Med. Internet Res. 22 (7), e17967 https://doi.org/10.2196/17967.

Constantino, M., Arnkoff, D., Glass, C., Ametrano, R., & Smith, J. (2011). Expectations. J. Clin. Psychol., 67(2). READCUBE. doi:https://doi.org/10.1002/jclp.20754.

Core Team, R., 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. https://www.R-project.org/.

de Winter, J.C.F., Dodou, D., Wieringa, P.A., 2009. Exploratory factor analysis with small sample sizes. Multivar. Behav. Res. 44 (2), 147–181. https://doi.org/10.1080/00273170902794206.

Dennison, L., Morrison, L., Lloyd, S., Phillips, D., Stuart, B., Williams, S., Bradbury, K., Roderick, P., Murray, E., Michie, S., Little, P., Yardley, L., 2014. Does brief telephone support improve engagement with a web-based weight management intervention? Randomized controlled trial. J. Med. Internet Res. 16 (3), e95 https://doi.org/10.2196/jmir.3199.

Devilly, G., & Borkovec, T. (2000). Psychometric properties of the credibility/expectancy questionnaire. Psychometric Properties of the Credibility/Expectancy Questionnaire., 31(2). READCUBE. doi:https://doi.org/10.1016/S0005-7916(00)00012-4.

Donkin, L., Glozier, N., 2012. Motivators and motivations to persist with online psychological interventions: A qualitative study of treatment completers. J. Med. Internet Res. 14 (3), e91 https://doi.org/10.2196/jmir.2100.

Donkin, L., Christensen, H., Naismith, S.L., Neal, B., Hickie, I.B., Glozier, N., 2011. A systematic review of the impact of adherence on the effectiveness of e-therapies. J. Med. Internet Res. 13 (3), e52 https://doi.org/10.2196/jmir.1772.

Donkin, L., Hickie, I.B., Christensen, H., Naismith, S.L., Neal, B., Cockayne, N.L., Glozier, N., 2013. Rethinking the dose-response relationship between usage and outcome in an online intervention for depression: randomized controlled trial. J. Med. Internet Res. 15 (10) https://doi.org/10.2196/jmir.2771.

Firth, J., Torous, J., Nicholas, J., Carney, R., Pratap, A., Rosenbaum, S., Sarris, J., 2017. The efficacy of smartphone-based mental health interventions for depressive symptoms: A meta-analysis of randomized controlled trials. World Psychiatry 16 (3), 287–298. https://doi.org/10.1002/wps.20472.

Flückiger, C., Del Re, A. C., Wampold, B. E., & Horvath, A. O. (2018). The alliance in adult psychotherapy: a meta-analytic synthesis. Psychotherapy (Chic.), 55(4), 316–340. doi:https://doi.org/10.1037/pst0000172.

Fuhr, K., Schröder, J., Berger, T., Moritz, S., Meyer, B., Lutz, W., Hohagen, F., Hautzinger, M., Klein, J.P., 2018. The association between adherence and outcome in an internet intervention for depression. J. Affect. Disord. 229, 443–449. https://doi.org/10.1016/j.jad.2017.12.028.

Graham, A.K., Lattie, E.G., Mohr, D.C., 2019. Experimental therapeutics for digital mental health. JAMA Psychiat. https://doi.org/10.1001/jamapsychiatry.2019.2075.

Harrell Jr., F. E. (2020). Hmisc: Harrell Miscellaneous (4.4-0) [Computer software]https://CRAN.R-project.org/package=Hmisc.

Hatcher, R.L., Gillaspy, J.A., 2006. Development and validation of a revised short version of the working alliance inventory. Psychother. Res. 16 (1), 12–25. https://doi.org/10.1080/10503300500352500.

Henson, P., Peck, P., Torous, J., 2019. Considering the therapeutic alliance in digital mental health interventions. Harv. Rev. Psychiatry 27 (4), 268–273. https://doi.org/10.1097/HRP.0000000000000224.

Hermes, E.D., Lyon, A.R., Schueller, S.M., Glass, J.E., 2019. Measuring the implementation of behavioral intervention technologies: recharacterization of established outcomes. J. Med. Internet Res. 21 (1), e11752 https://doi.org/10.2196/11752.

Hollis, C., Falconer, C.J., Martin, J.L., Whittington, C., Stockton, S., Glazebrook, C., Davies, E.B., 2017. Annual research review: digital health interventions for children and young people with mental health problems - a systematic and meta-review. J. Child Psychol. Psychiatry Allied Discip. 58 (4), 474–503. https://doi.org/10.1111/jcpp.12663.

Horn, J.L., 1965. A rationale and test for the number of factors in factor analysis. Psychometrika 30 (2), 179–185. https://doi.org/10.1007/BF02289447.

Horvath, A.O., Greenberg, L.S., 1989. Development and validation of the working alliance inventory. J. Couns. Psychol. 36 (2), 223–233. https://doi.org/10.1037/0022-0167.36.2.223.

Josephine, K., Josefine, L., Philipp, D., David, E., Harald, B., 2017. Internet- and mobile-based depression interventions for people with diagnosed depression: a systematic review and meta-analysis. J. Affect. Disord. 223, 28–40. https://doi.org/10.1016/j.jad.2017.07.021.

Kelders, S.M., Kip, H., Greeff, J., 2020a. Psychometric evaluation of the TWente engagement with Ehealth Technologies Scale (TWEETS): evaluation study. J. Med. Internet Res. 22 (10), e17757 https://doi.org/10.2196/17757.

Kelders, S. M., van Zyl, L. E., & Ludden, G. D. S. (2020b). The concept and components of engagement in different domains applied to ehealth: a systematic scoping review. *Front. Psychol.*, 11. doi:https://doi.org/10.3389/fpsyg.2020.00926.

Kroenke, K., Spitzer, R.L., Williams, J.B.W., 2001. The PHQ-9: validity of a brief depression severity measure. J. Gen. Intern. Med. 16 (9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x.

Lerner, J.S., Tetlock, P.E., 1999. Accounting for the effects of accountability. Psychol. Bull. 125 (2), 255–275. https://doi.org/10.1037/0033-2909.125.2.255.

Linardon, J., Cuijpers, P., Carlbring, P., Messer, M., Fuller-Tyszkiewicz, M., 2019. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 18 (3), 325–336. https://doi.org/10.1002/wps.20673.

Mohr, D.C., Duffecy, J., Jin, L., Ludman, E.J., Lewis, A., Begale, M., Jr, M.M., 2010. Multimodal E-mental health treatment for depression: a feasibility trial. J. Med. Internet Res. 12 (5), e48 https://doi.org/10.2196/jmir.1370.

Mohr, D.C., Cuijpers, P., Lehman, K., 2011. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J. Med. Internet Res. 13 (1), e30 https://doi.org/10.2196/jmir.1602.

Mohr, D.C., Lattie, E.G., Tomasino, K.N., Kwasny, M.J., Kaiser, S.M., Gray, E.L., Alam, N., Jordan, N., Schueller, S.M., 2019a. A randomized noninferiority trial evaluating remotely-delivered stepped care for depression using internet cognitive behavioral therapy (CBT) and telephone CBT. Behav. Res. Ther. 123, 103485. https://doi.org/10.1016/j.brat.2019.103485.

Mohr, D.C., Schueller, S.M., Tomasino, K.N., Kaiser, S.M., Alam, N., Karr, C., Vergara, J. L., Gray, E.L., Kwasny, M.J., Lattie, E.G., 2019b. Comparison of the effects of coaching and receipt of app recommendations on depression, anxiety, and engagement in the IntelliCare platform: factorial randomized controlled trial. J. Med. Internet Res. 21 (8), e13609 https://doi.org/10.2196/13609.

Newman, M.G., Fisher, A.J., 2010. Expectancy/credibility change as a mediator of cognitive behavioral therapy for generalized anxiety disorder: mechanism of action or proxy for symptom change? Int. J. Cogn. Ther. 3 (3), 245–261. https://doi.org/10.1521/ijct.2010.3.3.245.

Pelletier, L.G., Tuson, K.M., Haddad, N.K., 1997. Client motivation for therapy scale: a measure of intrinsic motivation, extrinsic motivation, and amotivation for therapy. J. Pers. Assess. 68 (2), 414–435. https://doi.org/10.1207/s15327752jpa6802_11.

Pihlaja, S., Stenberg, J.-H., Joutsenniemi, K., Mehik, H., Ritola, V., Joffe, G., 2018. Therapeutic alliance in guided internet therapy programs for depression and anxiety disorders—a systematic review. Internet Interv. 11, 1–10. https://doi.org/10.1016/j.invent.2017.11.005.

Pilutti, L., Dlugonski, D., Sandroff, B., Klaren, R., Motl, R., 2014. Randomized controlled trial of a behavioral intervention targeting symptoms and physical activity in multiple sclerosis. Mult. Scler. J. 20 (5), 594–601. https://doi.org/10.1177/1352458513503391.

Revelle, W. (2020). psych: Procedures for Psychological, Psychometric, and Personality Research (1.9.12.31) [Computer software]. https://CRAN.R-project.org/package=psych.

Rush, A.J., Bernstein, I.H., Trivedi, M.H., Carmody, T.J., Wisniewski, S., Mundt, J.C., Shores-Wilson, K., Biggs, M.M., Woo, A., Nierenberg, A.A., Fava, M., 2006. An evaluation of the quick inventory of depressive symptomatology and the Hamilton rating scale for depression: a sequenced treatment alternatives to relieve depression trial report. Biol. Psychiatry 59 (6), 493–501. https://doi.org/10.1016/j.biopsych.2005.08.022.

Schueller, S., & Mohr, D. (2015). Initial field trial of a coach-supported web-based depression treatment. Proceedings of the 9th International Conference on Pervasive Computing Technologies for Healthcare. 9th International Conference on Pervasive Computing Technologies for Healthcare, Istanbul, Turkey. doi:https://doi.org/10.4108/icst.pervasivehealth.2015.260115.

Schueller, S.M., Torous, J., 2020. Scaling evidence-based treatments through digital mental health. Am. Psychol. 75 (8), 1093–1104. https://doi.org/10.1037/amp0000654.

Schueller, S.M., Tomasino, K.N., Mohr, D.C., 2017. Integrating human support into behavioral intervention technologies: the efficiency model of support. Clin. Psychol. Sci. Pract. 24 (1), 27–45. https://doi.org/10.1111/cpsp.12173.

Sheehan, D.V., Lecrubier, Y., Sheehan, K.H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., Dunbar, G.C., 1998. The mini-international neuropsychiatric interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J. Clin. Psychiatry 59 (Suppl. 20), 22–33.

Sieverink, F., Kelders, S.M., van Gemert-Pijnen, J.E., 2017. Clarifying the concept of adherence to eHealth technology: systematic review on when usage becomes adherence. J. Med. Internet Res. 19 (12), e402 https://doi.org/10.2196/jmir.8578.

Smeets, R., Beelen, S., Goossens, M.E., Schouten, E.G., Knottnerus, J.A., Vlaeyen, J.W., 2008. Treatment expectancy and credibility are associated with the outcome of both physical and cognitive-behavioral treatment in chronic low back pain. Clin. J. Pain 24 (4), 305–315.

Sorkin, D.H., Janio, E.A., Eikey, E.V., Schneider, M., Davis, K., Schueller, S.M., Stadnick, N.A., Zheng, K., Neary, M., Safani, D., Mukamel, D.B., 2021. Rise in use of digital mental health tools and Technologies in the United States during the COVID-19 pandemic: survey study. J. Med. Internet Res. 23 (4), e26994 https://doi.org/10.2196/26994.

Stiles-Shields, C., Montague, E., Kwasny, M.J., Mohr, D.C., 2019. Behavioral and cognitive intervention strategies delivered via coached apps for depression: pilot trial. Psychol. Serv. 16 (2), 233–238. https://doi.org/10.1037/ser0000261.

Sucala, M., Schnur, J.B., Constantino, M.J., Miller, S.J., Brackman, E.H., Montgomery, G. H., 2012. The therapeutic relationship in e-therapy for mental health: a systematic review. J. Med. Internet Res. 14 (4), e110 https://doi.org/10.2196/jmir.2084.

Tomasino, K.N., Lattie, E.G., Wilson, R.E., Mohr, D.C., 2017. Coaching Manual for the ThinkFeelDo Internet Intervention Program. Northwestern University, Chicago, IL. https://doi.org/10.18131/G31T6P.

Torous, J., Lipschitz, J., Ng, M., Firth, J., 2020. Dropout rates in clinical trials of smartphone apps for depressive symptoms: a systematic review and meta-analysis. J. Affect. Disord. 263, 413–419. https://doi.org/10.1016/j.jad.2019.11.167.

Vogel, P.A., Hansen, B., Stiles, T.C., Götestam, K.G., 2006. Treatment motivation, treatment expectancy, and helping alliance as predictors of outcome in cognitive behavioral treatment of OCD. J. Behav. Ther. Exp. Psychiatry 37 (3), 247–255. https://doi.org/10.1016/j.jbtep.2005.12.001.

Wehmann, E., Köhnen, M., Härter, M., Liebherz, S., 2020. Therapeutic alliance in technology-based interventions for the treatment of depression: systematic review. J. Med. Internet Res. 22 (6), e17195 https://doi.org/10.2196/17195.

Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y., Zemla, J., 2017. *corrplot: Visualization of a Correlation Matrix* (0.84) [Computer software]. https://CRAN.R-project.org/package=corrplot.

Wickham, H. (2007). Reshaping data with the reshape package. Reshaping Data with the Reshape Package, 21(12). READCUBE. doi:https://doi.org/10.1016/S0142-1123(99)00007-9.

Wickham, H., 2009. ggplot2: Elegant Graphics for Data Analysis. Springer.

Wickham, H., Francois, R., 2015. *dplyr: A Grammar of Data Manipulation* (0.4.3) [computer software]. https://CRAN.R-project.org/package=dplyr.

Wickham, H., & RStudio. (2019). *tidyverse: Easily Install and Load the "Tidyverse"* (1.3.0) [Computer software]. https://CRAN.R-project.org/package=tidyverse.