

**Onsite Screening and Enhanced EAP Counseling Improves Overall Health, Depression and  
Work Outcomes: Four-Wave Longitudinal Pilot Study at Community Health Center in  
Vermont**

Mark Attridge, Ph.D., M.A.  
*Attridge Consulting, Inc.*  
*Minneapolis, Minnesota, USA*  
ORCID = [orcid.org/0000-0003-1852-2168](https://orcid.org/0000-0003-1852-2168)  
LinkedIn = <https://www.linkedin.com/in/markattridgephd>

Steven P. Dickens, M.A.  
*Invest EAP Centers for Wellbeing*  
*State of Vermont Division of Vocational Rehabilitation*  
*Burlington, Vermont, USA*

Author Contact Information:

Dr. Mark Attridge, President of Attridge Consulting, Inc., 1129 Cedar Lake Road South,  
Minneapolis, MN, USA. e-mail = [mark@attridgeconsulting.com](mailto:mark@attridgeconsulting.com)

**IN-PRESS**

**Journal of Workplace Behavioral Health: Employee Assistance Practice and Research**

**September 08, 2021**

### Abstract

Over a 2-year period, patients of a rural community health center in Vermont were screened onsite for multiple behavioral health risk factors and if found at-risk, were offered no-cost counseling onsite at the health center. The treatment was delivered by two licensed employee assistance program (EAP) counselors highly proficient in an evidence-based Motivational Interviewing approach. Longitudinal data at four time points were examined: baseline; end of treatment (3 months); at three- and six-months post treatment. Tests of paired data of baseline versus each later time point found significant improvement for global health (PROMIS-10) in the total sample ( $N = 120$ ); depression symptoms (PHQ-9) among subsample with depression as clinical issue ( $n = 68$ ); and work presenteeism and estimated hours of lost work time among employed subsample ( $n = 46$ ). Improvement in global health was significantly associated with improvement (reductions) in hours of lost work productivity at all later times ( $r = .33, .35, .50$ ). Improvement in global health was moderated by level of household income (more improvement among cases at lower income levels), but not by clinical or demographic factors. Practical implications for behavioral health screening and intervention (BSI) in medical settings and onsite EAP counseling programs are discussed.

KEY WORDS: EAP counseling, depression, global health, risk screening, work productivity

In this applied naturalistic study, we evaluate the outcomes of an expanded type of behavioral screening and intervention (BSI). Our intervention, while loosely based on Screening, Brief Intervention and Referral to Treatment (SBIRT), screened for a range of behavioral health risks, including substance use, depression, nutrition, exercise, smoking, and several more common issues seen in employee assistance programs (family and relationship problems, legal challenges, financial problems, stress, housing or transportation concerns, and childcare and eldercare matters). The clinical intervention averaged four treatment sessions (but included up to a dozen, as warranted by the presenting problem). Our clinical staff consisted of two highly trained employee assistance program (EAP) counselors, one a licensed mental health counselor and the other a licensed clinical social worker, both of whom received additional intensive training in Collaborative Care and Motivational Interviewing (MI). We provided the intervention onsite to adult patients at a community health clinic located in rural northeastern Vermont. This clinic was a Federally Qualified Health Center dedicated to providing primary care services to lower income patients. Our hypothesis guiding the project was that pairing an EAP counseling intervention with rigorous evidence-based behavioral health screening and interventions delivered onsite at a health care center would result in improved health and work outcomes.

Our BSI intervention is very different from the typical path to obtaining counseling from an EAP. First, in most EAPs, a passive approach is used in which the program waits for employees to initiate contact with the EAP when their personal situation worsens to the point where they feel professional help is needed. In contrast, in the BSI approach, we proactively screen an entire sub-population (which could be patients at a health care clinic – as in this study – or employees at a place of work) to identify individuals with behavioral risks. Second, in BSI,

the clinical treatment phase extends beyond the normal confines of an EAP intervention, employing Collaborative Care and Motivational Interviewing to address identified risk factors. The process often prompts individuals to identify issues they would like to work on with the counselor. Many of these individuals would likely have never contacted a counselor on their own. The two-step BSI process is therefore more proactive and more treatment-oriented than the traditional EAP.

### **Literature Review**

Behavioral health issues affect about one in every five adults (Karg et al., 2014). Numerous studies have demonstrated that behavioral health conditions – such as anxiety, depression, smoking, and drug use – are associated with increased overall health care costs (Goetzel et al., 2020; Kowlessar et al., 2011; O’Donnell et al., 2015). One study found that \$672 billion was spent per year in the US, on average, for adults treated for any type of behavioral health disorder during the years 2010–2013 (Thorpe et al., 2017). This high cost is driven in part by the comorbidity between behavioral health disorders and physical diseases (Milliman, Inc., 2020; Ornstein et al., 2013).

More specifically, the behavioral health disorder of depression affects about one in every 12 adults (8%; Karg et al., 2014) each year. Depression is also among the most burdensome of behavioral health disorders worldwide, giving rise to considerable adverse effects on activities of daily living and work for extended periods of time (Bruffaerts et al., 2012). Depression and other mental health problems are also typically among the most common reasons for use of brief counseling from employee assistance programs (EAP) sponsored by employers as an employee benefit (Attridge, 2019; Morneau Shepell, 2020). Therefore, it makes sense for employers to try

to prevent or reduce the rates of depression and other common behavioral health conditions among employees and family members.

Common behavioral health risk factors (particularly depression; see Berndt et al., 1998; Morneau Shepell, 2020) are associated with work performance risks such as employee absenteeism and presenteeism (Boles et al., 2004; Kowlessar et al., 2011; Mitchell & Bates, 2011; Mitchell et al., 2013; Shi et al., 2013). Two past research reports with data specific to EAP use also document the co-occurrence between improvements over time after use of counseling in areas of employee health and improvements in employee work performance (Attridge, 2016; Attridge et al., 2001). The relatively greater impact that behavioral health issues have on work presenteeism than on work absence also is a theme of past research (Lohaus & Habermann, 2019; Morneau Shepell, 2020). This link between behavioral health and work outcomes could be explored further in applied settings.

### **Behavioral Health Screening and Intervention**

Behavioral Screening and Intervention (BSI) is a method for screening and treating individuals for common behavioral health problems, such as depression, substance use, or smoking. Some research documents how participation in such programs can result in improved health and work outcomes and potentially reduced costs (Agerwala & McCance-Katz, 2012; Bray et al., 2007; Brown et al., 2014; Madras et al., 2009; White et al., 2013; White et al., 2014). Therefore, adding a BSI approach to standard approaches supporting workplace mental health from EAPs or other sources could help to reduce these risks. Providing the intervention services onsite after a screening at locations where many people are available and when they are already actively interested in their health could also potentially identify more of those at-risk and provide a prevention-oriented brief intervention to a larger segment of the population who are at-risk.

## **Overview of the Research Project**

Invest EAP Centers for Wellbeing is a public, not-for-profit, organization that operates within the State of Vermont's Division of Vocational Rehabilitation. The program provides an employee assistance program and other services to a broad cross section of the public and private sectors in Vermont. Approximately 20 percent of the state's entire population is covered by this organization. The Vermont Healthcare Innovation Healthcare Innovation Project – funded through Vermont's State Innovation Model (SIM) grant, awarded by the Center for Medicare and Medicaid Services (CMS) Innovation Center – provided a grant to evaluate the effectiveness of an innovative intervention combining aspects of behavioral screening, motivational interviewing, collaborative care and EAP.

This project site was a Federally Qualified Healthcare Center, Northern Counties Health Care, located in St. Johnsbury, Vermont. This center is less than 50 miles from the Canadian border and serves a mostly rural community of around 7,000 people. Most patients came to the center for other health care needs and as part of the routine check-in process were provided a brief 2-page screening to complete when they visited the health center. The responses to the screening questionnaire were scored immediately by staff at the center, and if the patient scored above the cutoff level for one or more of the risk measures, then he or she was invited to participate voluntarily in further counseling provided weekly at the health center.

The funding associated with the research grant limited the total number of cases allowed to participate in the study to 120 people. Depending on the level of interest from the patient in getting health coaching and the limited availability of appointment times with the counselor, only some who were screened participated in the counseling while many others did not. Thus, we ended up with a much larger pool of people with data on the risk screener tool than the number

of people who participated in the counseling treatment part of the intervention. These non-participants were used as a comparison group for examination of the representativeness of the counseling participants concerning demographic characteristics. To further inform this normative comparison, we also identified external benchmarks on some demographic factors for adults in Vermont from the literature.

### **Research Questions**

The focus of this paper was to answer questions regarding the general effectiveness of our BSI approach and investigate if there was a link between health and work outcomes. More specifically:

RQ1: Is the counseling effective in improving global health?

RQ2: Are any of the client demographic or clinical delivery factors moderators of how much global health improves after counseling?

RQ3: Is the counseling effective in improving depression symptoms for the subsample of cases with depression as their clinical goal?

RQ4: For the employed cases, is the counseling effective in improving work presenteeism and hours of lost productive work time?

RQ5: For the employed cases, are improvements in global health associated with improvements in hours of lost productive work time?

## **Methods**

### **Risk Screening for Behavioral Health Risks**

An initial risk screening was done to qualify individuals to be eligible to participate in the study and receive the counseling intervention. The two-page screening instrument contained 14 items (involving 22 distinct responses). It was created by the authors in collaboration with Dr.

Richard Brown (2016) of the University of Wisconsin School of Medicine and Public Health. The screening items identified risks in the areas of diet/nutrition, physical exercise, depression, smoking, alcohol, drugs, nuisance health problems, and personal life concerns. It also had three items related to work, including presenteeism for daily life tasks or work, performance level for daily life tasks or work, and hours of work absence (if employed). The screening was given in paper format to patients at the health clinic who were there for routine or emergent medical services. The patient's responses to the screening form were scored immediately by clinic staff, and those who scored above the cutoff level for one or more risk areas were invited to participate in the counseling study.

### **Study Design for the Counseling Intervention**

The intervention part of the project involved a non-experimental, prospective, longitudinal research study design with a single group who experienced a counseling intervention. The nature and length of the counseling provided varied depending on the unique needs of each client. The group that received the BSI counseling intervention had repeated measurement of multiple outcomes at four points in time, including:

- Time 1 at baseline (i.e., at the first clinical session)
- Time 2 at the end of treatment (i.e., final scheduled clinical session)
- Time 3 at first follow-up at 3-months post treatment
- Time 4 at second follow-up at 6-months post treatment

The targeted amount of time elapsed between each phase of data collection planned in the study design was achieved on average, with 3.7 months between the end of treatment and the first follow-up (versus 3.0 planned) and 6.8 months between the end of treatment and the final follow-up (versus 6.0 planned). See Table 1 for details on this factor. The screener and first



session were done separately with screener data not matched up to case's same Time 1 questionnaire responses. The Time 1 assessment was done onsite in paper format before the counseling started whereas each of the later points had questionnaires completed online via email invitation to a secure website that used a unique identification code for each patient.

### **Clinical Participation**

Entry into the counseling intervention phase for specific participants was not randomized or assigned in a systematic experimental fashion. Rather, the participation in counseling was voluntary and based on a combination of factors, including timing, such as when the primary care physician needed to meet with the patient and when the onsite counselor was free. All counseling services were provided in face-to-face sessions delivered onsite at the community health center. The clinical contact occurred for under one hour per session (average of 37 minutes). The typical case engaged in four sessions of counseling over a three-month treatment period (or about one session per every three weeks).

### **Clinical Providers**

The two counselors who provided the clinical interventions were a master's level Licensed Clinical Mental Health Counselor and a Licensed Clinical Social Worker. Both were senior-level counselors with over 20 years of professional clinical experience. Prior to the intervention phase of the study, these staff participated in a weeklong intensive training offered by faculty from the University of Wisconsin School of Medicine and Public Health and a Motivational Interviewing Network of Trainers (MINT) certified trainer that focused on Motivational Interviewing and Collaborative Care. The staff then participated in regular telephone supervision with these trainers for a 6-month period. The counselors were required to submit tape recordings of sessions and to be scored for proficiency in Motivational Interviewing

by the MINT trainer. Counseling staff also participated in the webinar-based training in Collaborative Care and Problem-Solving Therapy offered by the University of Washington.

### **Time Period**

The time frame for the counseling experience and follow-ups was from the latter part of year 2015 through the entire year of 2016.

### **Samples**

**Treatment Group - Total.** Self-report data were collected from 120 participants who completed the counseling intervention and had both baseline data and outcome data from at least one or more of the later three time points. The profile for these cases is presented later in the paper.

**Treatment Group – Employed.** Although 54 of the original 120 cases reported working at baseline, over the course of the study period, the size of the employed group changed. There were 46 cases that were continuously employed, eight cases that were employed but changed later became unemployed, and another 14 cases that were initially unemployed and became employed after counseling. This consistently employed subsample of 46 cases was 89% White race, 65% female, 63% married, 59% had one or more children living at home and had an average age of 43 years. Employed cases had an average of 4.2 counseling sessions during the intervention phase. These characteristics for the employed subsample were similar to the full treatment group.

**Normative Data on Demographic Factors.** During the same period, the community health center also had over 2,000 other patients who took the same BSI screener questionnaire (also via a paper format while waiting for health care services) but who had no sessions with the onsite counselor. At the time, the raw data from these screening forms was not coded for

analysis. However, a sample of 500 individuals who completed the screener questionnaire but who did not see the counselor were randomly selected on a post hoc basis (using a statistical procedure in SPSS) to form a normative comparison group for the counseling participants. A total of 328 people in the comparison group were employed. In general, the background characteristics of this employed subsample were similar to the overall non-treatment sample. Certain demographic norms for the state of Vermont were also obtained from the Centers for Disease Control's Behavioral Risk Factor Surveillance System (BRFSS) report, which examined data from epidemiological random sample interviews conducted in 2016 ( $N > 3,500$ ; Vermont Department of Health, 2017).

### **Representativeness of the Treatment Group**

To determine whether the self-selected treatment group was representative of the larger local population, the profile of the counseling treatment participants at baseline was compared against 1) other people from the same community health center who did not participate in the counseling and 2) from other external normative data. The treatment group was similar to the state norms for the background characteristics of race (both 94% White) and marital status (48% vs. 53%), but there were statistically significant differences (all  $p < .05$ ) on other personal attributes. Average age differed between the three groups ( $t = -2.23$ ;  $t = -9.72$ ), but it was a small difference in a practical sense with only 4 to 6 years difference (treatment = 46 years; non-treatment norm = 52 years; and state norm = 48 years). Almost three-fourths of the treatment group were women (73%), compared to 65% of the non-treatment sample and 51% statewide ( $X^2 = 30.82$ ). Despite similar age and marital status profiles, the treatment group had a higher percentage of people with one or more children living at home (42%) compared to the statewide average of 30% ( $X^2 = 6.03$ ). The treatment group also had far more cases who were unemployed

(62%) compared to both the non-treatment group norm (34%) and the statewide norm (38%;  $\chi^2 = 31.43$ ). The treatment group also had a much lower income profile compared to the state norms ( $\chi^2 = 85.90$ ): annual household income under \$25,000 per year was 66% of the treatment cases versus just 25% norm; \$25,000 to under \$50,000 was 23% treatment vs. 25% norm; \$50,000 to under \$75,000 was 8% treatment vs. 18% norm; and \$75,000 or more was 3% treatment vs. 32% norm. To summarize, comparison of the treatment group with a random sample of non-participants at the same health clinic and with the state norms indicated the treatment group had a similar profile on background characteristics of age, race, and marital status but were somewhat different on gender (more women), employment status (more were unemployed), and income level (far more of lower income). Given that the mandate of this publicly funded clinic is to provide health care to underserved populations, the lower income profile of the study participants is consistent with the larger patient community at this setting.

## Measures

This report analyzes two health and four work-related outcomes. All of these measures had acceptable levels of internal reliability (for multi-item scales) and test-retest stability, despite expecting change in scores post treatment which appropriately reduces the test-retest association.

**Global Health.** The 10 items of the Global Health assessment component of the Patient-Reported Outcomes Measurement Information System (PROMIS) questionnaires were used to assess overall health (Hays et al., 2009). PROMIS is a National Institutes of Health (NIH) Roadmap initiative in the United States designed to improve self-reported outcomes using state-of-the-art psychometric methods. Adapting the World Health Organization's tripartite framework of physical, mental, and social health, the PROMIS measures were developed and calibrated on very large national samples. Global health items are evaluations of health in

general rather than specific elements of health. These items allow an efficient assessment of self-reported general health. The PROMIS Scales have established measurement reliability and validity (Cella, et al., 2010; Magasi, et al., 2011; PROMIS, 2017). These items are predictive of important future events such as health care utilization and expenditures and mortality (Bjorner, et al., 2005).

*PROMIS-1 General Health Status.* Item = *In general, would you say your health is:*

Response options of: Excellent (5), Very Good (4), Good (3), Fair (2) or Poor (1).

*PROMIS-2 Physical Health Status.* Item = *In general, how would you rate your physical health?* With 5-point response scale of: Excellent (5), Very Good (4), Good (3), Fair (2) or Poor (1).

*PROMIS-3 Physical Activities of Daily Living.* Item = *To what extent are you able to carry out your everyday physical activities such as walking, climbing stairs, carrying groceries, or moving a chair?* With 5-point response scale of: Completely (5), Mostly (4), Moderately (3), A little (2), or Not at all (1).

*PROMIS-4 Pain.* Item = *In the past 7 days, how would you rate your pain on average?* With an 11-point response scale of: No Pain = 0 1 2 3 4 5 6 7 8 9 10 = Worst imaginable pain. Following Hays et al. (2009), we recoded the pain intensity item from the 0–10 rating scale to 5 categories (similar to the other PROMIS items) based on grouping of 0–10 response options as follows: 0 = 1; 1–3 = 2; 4–6 = 3; 7–9 = 4; 10 = 5.

*PROMIS-5 Fatigue.* Item = *In the past 7 days, how would you rate your fatigue on average?* With 5-point response scale of: None (1), Mild (2), Moderate (3), Severe (4), and Very severe (5).

*PROMIS-6 Mental Health Status.* Item = *In general, how would you rate your mental*

*health, including your mood and your ability to think?* With 5-point response scale of: Excellent (5), Very Good (4), Good (3), Fair (2) or Poor (1).

*PROMIS-7 Emotional Distress.* Item = *In the past 7 days, how often have you been bothered by emotional problems such as feeling anxious, depressed or irritable?* With 5-point response scale of: Never (5), Rarely (4), Sometimes (3), Often (2), or Always (1).

*PROMIS-8 Quality of Life.* Item = *In general, would you say your quality of life is:* With 5-point response scale of: Excellent (5), Very Good (4), Good (3), Fair (2) or Poor (1).

*PROMIS-9 Social Satisfaction.* Item = *In general, how would you rate your satisfaction with your social activities and relationships?* With 5-point response scale of: Excellent (5), Very Good (4), Good (3), Fair (2) or Poor (1).

*PROMIS-10 Social Activity.* Item = *In general, please rate how well you carry out your usual social activities and roles. (This includes your activities at home, at work and in your community, and your responsibilities as a parent, child, spouse, employee, friend, etc.).* With 5-point response scale of: Excellent (5), Very Good (4), Good (3), Fair (2) or Poor (1).

*Summary Scale Psychometrics.* A PROMIS-10 summary scale was created using all 10 items. The items on pain, fatigue and emotional distress had to be reverse scored to be included in the scale. Higher scores on this scale indicate better health with a range of 10 to 50. This measure had good internal reliability: Time 1  $\alpha = .84$  ( $n = 118$ ); Time 2  $\alpha = .86$  ( $n = 84$ ); Time 3  $\alpha = .80$  ( $n = 86$ ); and Time 4  $\alpha = .91$  ( $n = 49$ ). The test-retest stability for the scale was acceptable: Time 1 with Time 2,  $r = .55$ ; Time 1 with Time 3,  $r = .61$ ; and Time 1 with Time 4,  $r = .38$ .

**Depression.** Depression was assessed with the Patient Health Questionnaire 9-item scale (PHQ-9). This scale has been used in hundreds of research studies and has established validity

and reliability (Kroenke & Spitzer, 2002; Martin et al., 2006). The instructions state: **Over the last 2 weeks**, how often have you been bothered by any of the following problems? Response options were (1) Not at all; (2) Several days; (3) More than half the days; and (4) Nearly every day. The items include: (a) *Little interest or pleasure in doing things*; (b) *Feeling down, depressed or hopeless*; (c) *Trouble falling or staying asleep, or sleeping too much*; (d) *Feeling tired or having little energy*; (e) *Poor appetite or overeating*; (f) *Feeling bad about yourself -- or that you are a failure or have let yourself or your family down*; (g) *Trouble concentrating on things, such as reading the newspaper or watching television*; (h) *Moving or speaking so slowly that other people could have noticed -- Or the opposite, being so fidgety or restless that you have been moving around a lot more than usual*; and (i) *Thoughts that you would be better off dead or hurting yourself in some way*. The PHQ-9 is scored by adding together the 9 items. Higher scores on this measure indicate greater depression. Scores are generally graded on level of severity, with minimal depression (0-4), mild depression (5-9), moderate depression (10-14), moderately severe depression (15-19), and severe depression (20-27). According to Kroenke (2012), a score of 10 or more is considered positive risk status for depression. This measure had good internal reliability: Time 1  $\alpha = .88$  ( $n = 112$ ); Time 2  $\alpha = .89$  ( $n = 84$ ); Time 3  $\alpha = .87$  ( $n = 87$ ); and Time 4  $\alpha = .93$  ( $n = 49$ ). The test-retest stability for this scale was acceptable: Time 1 with Time 2,  $r = .55$ ; Time 1 with Time 3,  $r = .55$ ; and Time 1 with Time 4,  $r = .62$ .

**Work Productivity.** A single item was adapted for this study from the job performance question from the World Health Organization's Health and Productivity Questionnaire (HPQ; Kessler et al., 2003). Their original item is "On a scale from 0 to 10 where 0 is the work job performance anyone could have and 10 is the performance of a top worker, how would you rate your overall job performance on the days you worked during the past 4 weeks (28 days)? The

item in our study asked: *During the past 4 weeks, how would you rate your overall ability to perform daily tasks and be productive at work or home given any life issues that may have impacted your focus or motivation? Please use the rating scale of 0 to 10, where 0 is the worst performance and 10 is the top performance = 0 1 2 3 4 5 6 7 8 9 10.* The test-retest stability for this item was modest, but acceptable considering we expected some change in the score (instability) over time after treatment: Time 1 with Time 2,  $r = .49$  ( $n = 82$ ); Time 1 with Time 3,  $r = .33$  ( $n = 85$ ); and Time 1 with Time 4,  $r = .25$  ( $n = 47$ ).

**Work Presenteeism.** A single-item was adapted for this study from the single-item presenteeism question from the Workplace Outcome Suite (WOS, Lennox et al., 2018). The WOS presenteeism item is “My personal problems kept me from concentrating on my work” for past 30 days and rated on 1-5 Likert type response scale of strongly disagree to strongly agree. Our adapted item asked: *How often during the past 4 weeks did health issues or dealing with life keep you from focusing fully on your work or daily tasks?* This was rated on a 1-5 scale of frequency: (1) None, (2) A little; (3) Some of the time; (4) Most of the time; and (5) All of the time. Higher scores indicate greater presenteeism. The test-retest stability for this item was modest: Time 1 with Time 2,  $r = .28$  ( $n = 26$ ); Time 1 with Time 3,  $r = .26$  ( $n = 32$ ); and Time 1 with Time 4,  $r = .29$  ( $n = 17$ ). On the original item from the WOS, having a “problem level” of work presenteeism is considered when the employee agrees with question (rating of 4 or 5 = problem; other ratings 1-3 = No problem; see Attridge et al., 2018). In this study, we considered a “problem” level of presenteeism to be when lack of work focus occurred “most of the time” or “all of the time” (ratings of 4 or 5).

**Work Absence Hours.** Absence from work was measured by an adapted version of the single-item absenteeism question from the Workplace Outcome Suite (WOS, Lennox et al.,



2018). The original item asks “*For the period of the past 30 days, please fill in the total number of work **hours** your personal concern caused you to be miss work. Include complete eight-hour days and partial days when you came in late or left early*” with a fill in the blank response. In this study we asked: *If you work, **during the past 4 weeks**, how often did health issues or dealing with life problems (such as the above list of concerns) cause you to be late to work, to leave work early or to miss a full day of work? Please fill in the total number of work **hours** missed (use a whole number, i.e., 5).* This measure uses the number of hours provided, with a maximum capped at 160 hours. Higher scores indicate greater absence from work. As in most other studies of work absence (Morneau Shepell, 2020), this measure was highly skewed, with the majority of cases reporting zero hours of absence at each time point (i.e., Time 1 = 52% of employees had zero absence; Time 2 = 62% zero; Time 3 = 53% zero; and Time 4 = 88% zero). The test-retest stability for hours of absence was: Time 1 with Time 2,  $r = .76$  ( $n = 26$ ); Time 1 with Time 3,  $r = .10$  ( $n = 32$ ); and Time 1 with Time 4,  $r = .52$  ( $n = 17$ ).

**Work Lost Productive Time.** The hours of work absence and work productivity level rating were combined into a single metric of lost productive time (LPT; see Stewart et al., 2003, American Productivity Audit studies). This metric starts with considering the total hours scheduled to work in a month. We assume a 160-hour standard full-time schedule. We did not ask how many hours the person was scheduled to work. From the scheduled total, we first deduct the hours lost to absence from work. From a 160-hour schedule, we deduct the hours of absence at baseline (assume this is 10 hours), which leaves 150 hours that were actually worked. Next, we use the work productivity 0-10 rating results to determine how much of the remaining time at work was unproductive time. The 0-10 rating reflects the full range of low to high work productivity and when multiplied by 10, it becomes a metric of 0% to 100% of work time.

Assume that at baseline there was a 60% level of work productivity. The amount of unproductivity is the difference between this level and the maximum of 100%. In this example, 40% of the time worked was unproductive (100% maximum minus the 60% at baseline). To get the number of hours of unproductive time, we multiply the actual hours worked of 150 by the 40% level of unproductivity. This amount of unproductive time (60 hours) is then combined with the number of hours of absence from work (10) to yield the total Lost Productive Time (LPT) result (i.e.,  $60 + 10 = 70$  hours). This LPT metric was examined for how much it changed from before to after the use of the counseling intervention. This analytical approach has been used successfully in other applied studies of the workplace outcomes from EAP counseling (Attridge, 2012, 2015, 2016; Morneau Shepell, 2020). The test-retest stability for this measure was modest: Time 1 with Time 2,  $r = .61$  ( $n = 26$ ); Time 1 with Time 3,  $r = .32$  ( $n = 32$ ); and Time 1 with Time 4,  $r = .36$  ( $n = 17$ ).

### **Statistical Analysis Plans**

Data were missing data for some cases at each of the three points after baseline. Therefore, a repeated measures longitudinal analysis involving the data at all four time point from every participant was not feasible – as this applied to only 27 of the 120 total cases. An alternative post hoc testing approach utilized paired tests of outcomes conducted for subsamples of the treatment cases having data at the start of treatment and also at a later point. More specifically, the maximum sample sizes for repeated measures pairings included: Time 1 paired with Time 2 had 84 cases; Time 1 paired with Time 3 had 85 cases; and Time 1 paired with Time 4 had 49 cases. However, some measures had missing data from a small number of cases and thus the actual sample sizes in certain specific tests had fewer cases (see results presented later).

The descriptive profile of the full sample at baseline and each of the three later pairings of baseline with a later time point are shown in Table 2. The seven demographic client factors and four clinical use factors were quite similar (most factors had a 5% or less difference between the three pairings). This similarity indicates that the paired testing approach had similar profiles of the measured background and treatment factors in each pairing, even though the actual cases represented in each pairing were a mix of the same or different specific clients.

Multivariate repeated measures analyses were also conducted to explore the possible influence of various client background characteristics and clinical experience factors on the extent of change in global health after treatment. These moderator tests were conducted only in the Time 1 and Time 3 pairing sample as this pairing had the highest number of cases. Each demographic and clinical factor was first tested independently. A final multi-variate model analysis was planned to examine the group of factors that each were significant as moderators when first tested separately.

Correlations were also obtained between standardized difference scores of the global health and lost work time measures to examine the possible relationship between these two outcomes. Difference scores were created for the change health score and change in the LPT hours at three periods: From baseline to end of treatment (test condition 1); from baseline to the 3-month follow-up (condition 2); and from baseline to the 6-month follow-up (condition 3). Each difference score was constructed to represent improvement as scores increased from low to high. For example, the PROMIS-10 summary score for Time 2 was subtracted from the PROMIS-10 summary score for Time 1, which yielded a difference such that a range of positive and negative scores was possible, with a positive score indicating improvement global health status, a score of zero indicating no change, and a negative score indicating a decline in global

health status. To reduce the skew in each distribution and correct for the differences in range of these differences scores, each measure was then converted to standardized z-scores so that the two improvement outcomes at each time pairing all had the same score distribution (near normal), general range, mean of zero, and standard deviation of 1.0.

Most tests used the statistical significance probability threshold of  $p < .05$  to be declared “significant”, but in the smaller sample size test conditions (i.e.,  $< 30$  cases) we allowed a  $p < .10$  result as “significant.” The Cohen  $d$  statistical effect size estimate also was calculated for the repeated measures analyses (Cohen, 1988). We used cut-offs of  $d = .20$  for small size,  $.50$  for medium size and  $.80$  or greater for large size to interpret the magnitude of effects.

### **Research Ethics**

The project was reviewed and approved by both the funding agency and the health center. The counselors who provided the treatments were licensed professionals. All participants signed consent forms. Questionnaire data were collected using unique identification numbers to protect the identity of individual participants.

## **Results**

### **Part 1: Improvement After Counseling in Global Health**

This section explored if overall health for clients in the treatment group was improved after experiencing the clinical intervention and over the two longitudinal follow-ups. The scores for the counseling cases on the global health measure are shown in Table 3 for each of the four time periods (with different sample sizes). Examination of Table 3 reveals a consistent pattern of results such that the average score for global health was lowest at baseline, higher at the end of treatment and then stayed higher at both follow-ups. The paired  $t$ -tests comparing the start of treatment to each of the later time points all were statistically significant ( $p < .001$ ) and each was

also a large size statistical effect. On average, the score for global health was improved by 21% at the end of treatment compared to baseline, by 16% at the 3-month follow-up compared to baseline, and by 29% at the 6-month follow-up compared to baseline.

## **Part 2: Moderators of Improvement After Counseling in Global Health**

We also explored if the extent of improvement experienced in global health after counseling was moderated by any of the demographic or clinical delivery factors. The top of Table 4 shows the results of repeated-measures ANOVA tests conducted separately with each of the six demographic factors and the three clinical factors. Background factors of client age, marital status, a child living at home, and employment status all were non-significant ( $p > .05$ ) as moderators of the extent of improvement in global health after counseling. However, household income ( $p < .01$ ) did moderate the extent of improvement in global health after counseling at the 3-month follow-up. More specifically, the most improvement in health was found among those cases with low income (24% average improvement in the PROMIS-10 score compared to the start of treatment) or cases with moderate income (17% improvement) and yet almost no change in global health among those cases with high income (1% improvement). Gender of the client also had an effect that approached significance ( $p = .09$ ), with females experiencing a 19% average improvement in global health scores compared to a 10% average for males.

For the clinical factors, completing the final planned counseling session (or not) and the topic of clinical focus during treatment (five types) both did not significantly influence the extent of improvement in global health. However, total number of sessions of counseling experienced did positively moderate the extent of improvement in global health after counseling ( $p < .05$ ). More specifically, the most improvement in health was found among those cases with the highest number of sessions used (6 to 10 sessions) who had a 21% average improvement in the

PROMIS-10 score compared to the start of treatment. The cases with an average number of sessions (3 to 5 sessions) had an average of 16% improvement. The least improvement in health (9% improvement) was found among those cases who used only 1 or 2 sessions.

When the set of the above three factors with effects was included in a multivariate analysis, only the interaction effect of income level by time remained significant (see bottom of Table 4), whereas the number of clinical sessions and gender were not significant anymore when also considering the stronger influence of income level. A closer inspection of the results shows that the cases in both the low- and middle-income groups had a baseline level of health status which was significantly worse than cases in the higher income group.

Indeed, the starting level for the higher income group was above post-treatment level of the other groups. This subgroup difference in average overall level of health by income level was significant (see between-subjects effect at bottom of Table 4). The final interpretation is that (lower) income was the only background or clinical factor to influence the general finding of greater improvement in global health after counseling, when also considering other relevant factors.

### **Part 3: Improvement After Counseling in Depression Symptoms Among Subsample of Cases Focused on Issue of Depression**

The above positive results obtained in the total study sample focused on the outcome of global health symptoms that spanned physical, mental, and social domains. Yet these cases had a range of different clinical issues that were focused on during treatment. Getting such a diversity of issues during treatment for different cases was precisely the point of using a multi-domain risk assessment screening at the start of the BSI process. The test results shown in Table 4 indicated that the average extent of improvement in PROMIS-10 scores from the start of

treatment to the 3-month follow-up was *not* significantly different between the various clinical issues. Nonetheless, there was a range in level of success observed by issue, with an average of 17% improvement for the subgroup of cases with depression issue, 7% improvement for the subgroup with smoking issue, 8% improvement for the subgroup with drinking issue, 13% improvement for the subgroup with wellness issues of diet or exercise and 32% improvement for the subgroup with other issues. These differences were not significant in the moderator tests, in large part because of the small numbers of cases within each subgroup (i.e., only 5 to 14 cases per each subgroup) and the associated higher variances around the respective mean scores. In contrast, the depression issue had 51 of the 85 total cases in that moderator test.

Of the different clinical issues represented in this study of people identified based on screening for multiple kinds of behavioral health risks, only one issue – depression – had enough cases to justify conducting a statistical examination of change over time in that subsample for an outcome matched to the clinical issue. By comparison, each the other clinical issues had fewer than the minimum of 30 cases preferred for a reliable test. The depression outcomes in this subsample are shown in Table 5. The first metric examined was the percentage of cases that had moderate or higher severity of depression symptoms (i.e., above the cut-off score on the PHQ-9 measure representing a “clinical” level of depression). This outcome was examined at each of the four time points. Having an “at-risk” status for depression severity changed from 74% of this subgroup before treatment, to 22% at the end of treatment, to 35% at the 3-month follow-up and to just 14% at the 6-month follow-up. Thus, about two-thirds of the cases that focused on depression issue during treatment moved from being at-risk at baseline to not at-risk anymore after treatment.

A more precise method of probability testing was conducted using the full range of the PHQ-9 scale (score 0-27). As shown in the lower part of Table 5, the paired *t*-tests at each time pairing all showed significant reductions ( $p < .001$ ) from the starting level of symptom severity compared to the later points after treatment. Each of these tests of the reduction in depression symptom severity scores after counseling had a large statistical effect. On average, when compared to baseline the score for depression symptom severity was reduced by 50% at the end of treatment, by 45% at the 3-month follow-up, and by 62% at the 6-month follow-up. The careful reader will note that improvement in depression scores among this depression-matched subgroup was about twice the average level of improvement (on a percentage basis above baseline) achieved on the global health outcome measure in the full sample across all cases regardless of clinical issues. More specifically, a 50% improvement in the depression outcome among cases who focused on depression during their treatment vs. 23% improvement on the global health outcome for all cases.

For the employed cases with a clinical issue of depression ( $n = 22$ ), the greater the severity of depression symptoms at baseline, the greater their work presenteeism at baseline ( $r = .40, p = .08$ ), the lower their work productivity at baseline ( $r = -.57, p < .01$ ), and the greater their combined hours of lost work productive time at baseline ( $r = .60, p < .01$ ). These findings replicate other studies in the literature (e.g., Shi et al., 2013) that link depression with various aspects of work performance. This set of correlations at baseline suggests that if the behavioral health issue can be improved after treatment, then perhaps work performance also can be improved. This idea is examined in the next part of the results.

#### **Part 4: Improvement After Counseling in Global Health and Work Outcomes Among Subsample of Employed Cases**



This section presents results for tests conducted within the employed subsample in the treatment group. Outcomes of global health, work presenteeism and hours of lost productive time are examined.

*Global Health Outcome.* The scores for the employed counseling cases on the global health measure are shown in the bottom of Table 3 for each of the four time periods of data collection (each with a different sample size). Examination of these findings reveals a consistent pattern of results such that the average score for global health was lowest at baseline, then higher at the end of treatment, and then stayed at the same higher level at both of the follow-ups. The paired *t*-tests comparing the start of treatment to each of the later time points all were statistically significant ( $p < .001$ ) and were either a medium or large statistical effect. Thus, these cases in the employed subsample improved after treatment compared to their baseline levels in the outcome of global health, ranging from 13% to 28% improvement (depending on time period).

*Work Presenteeism.* The improvement after counseling in the ability to fully focus on work is shown in Table 6. Experiencing presenteeism at a problem level of “most” or “all of the time” in the past month was found for one-third of the employed cases at baseline when starting treatment. This changed to only 15% of all employed cases at the end of treatment and was even lower at both follow-ups (3% and 6%). Statistical tests of the full 1-5 ratings for the work presenteeism item showed the same pattern (see Table 6), with the baseline average score being significantly higher (more presenteeism) than the averages at each of the three later points in time. These differences ranged from a 23% to 43% improvement from the baseline levels and were either medium or large size statistical effects. As with the health outcome, work presenteeism also improved from the baseline to the end of treatment, and improvement was maintained at both follow-ups.

Although work presenteeism and lost productive time are distinct measures, they are conceptually related as they represent primarily cognitive and behavioral manifestations of the same general problem. As such, it was no surprise that these two aspects of work impairment were positively correlated at each period: Time 1,  $r = .39$ , Time 2,  $r = .67$ ; Time 3,  $r = .65$ ; and Time 4,  $r = .51$ ; all significant at  $p < .05$ .

*Work LPT.* The hours of lost productive work time are shown in Table 7 for each of the four time periods. The results reveal a consistent pattern such that the average hours of LPT was highest at baseline (77 hours) but then changed to be lower at the end of treatment (50 hours) and continued to stay lower at both the 3-month (49 hours) and 6-month (40 hours) follow-ups. Tests comparing the start of treatment in pairings with available cases at each later point were all statistically significant ( $p < .01$ ) and were either a medium or a large statistical effect. On average, in these pairings, the number of estimated hours of lost productive work was reduced by 36% at the end of treatment compared to baseline, reduced by 33% at the 3-month follow-up compared to baseline, and reduced by 47% at the 6-month follow-up compared to baseline.

#### **Part 5: Co-occurrence of Improvement in Global Health Risks and Improvement in Work LPT Hours Among Subsample of Employed Cases**

We also examined the linkage between improvements after treatment in employee health and improvements in work through generating correlations of the standardized difference score measures for global health and LPT hours. As shown in Table 7, the level of improvement on work outcome of LPT hours was about twice the average level of improvement on global health – and this was true at each time pairing. Among the employed cases, the extent of improvement in health from the start of treatment to later points after treatment was positively associated with the extent of reduction in the hours of LPT during the same corresponding points in time (see

Table 8,  $r = .33, .35$  and  $.50$ ). Thus, improvement post treatment in global health was consistently associated with improvement post treatment in restored hours of work productivity (i.e., less absence and more productivity while on the job associated with improved overall health). Moreover, the size of the change was relatively greater for this composite work outcome than for the global health outcome (i.e., average percentage improvement after treatment of 47% for work LPT hours vs. 21% for global health).

### **Discussion**

We found that after brief psychological treatment provided onsite at a local health clinic, participants experienced significant improvements in both global health for all cases and for depression symptoms for cases who had a depression issue. Among the employed cases, both adverse outcomes of work presenteeism and lost productive hours were reduced by treatment. Most importantly, improvements achieved at the end of treatment for all of these outcomes were maintained at both 3- and 6-months later.

Moreover, we found evidence that improving the global health of employees was associated with also improving their work presenteeism and productivity. This last finding should be of special interest to employee assistance programs, which according to the Employee Assistance Professionals Association (2011), an “EAP is a workplace program designed to assist...work organizations in addressing productivity issues.” These findings offer correlational evidence that reducing clinical distress is associated with restored work performance. This finding is particularly interesting when none of the clients had work issues as the focus of the counseling intervention – and yet their work tended to get better over time as their health got better.

All of the counseling clients were volunteers who chose to participate in the study. Others, including some who screened positive for one or more behavioral health risks, chose not to avail themselves of the offer of counseling. While that may suggest the possibility of some selection bias, it should be kept in mind that this is exactly how any BSI program would likely operate in the real world. Screening and intervention would be offered to a broad population; only some would choose to participate, and an even smaller number who choose to participate and who screened positive for behavioral health risks would choose to engage in counseling aimed at facilitating behavioral change.

### **Study Limitations**

As this was a pilot project, it has several limitations that must be considered. It was conducted at a single site, with a relatively small sample size (especially for the employee subsample), and with declining levels of participation across the four-wave longitudinal study design. The comparison to local non-participants and to state norms, however, indicated the counseling participants were similar on most background factors to other norms and even the lower income status of most participants was consistent to the local patient population of the community health center.

The lack of full data for every case at all four time points did not allow us to use the preferred analytical approach of conducting repeated measures analyses with all four waves of data in the same test. The option to statistically estimate these missing scores, while possible to do using a computer program in SPSS, we judged as inappropriate because it would have involved estimating new data for too many of the cases at the later time points (30%; 28%, and 59% of cases that were missing at Time 2, Time 3, and Time 4). Thus, we settled on testing each

of the later points against the starting level at baseline. This approach did find the expected improvement results consistently over all three of the study time periods.

The lack of an experimental control group is a more serious limitation. It is unknown how much the prevalence rates for these risk factors would have changed over time without an intervention.

A final limitation was our reliance on self-report sources for the data. Other records or external sources of the outcome measures used in the study could have provided more accurate measurements for some outcomes. However, there is a large literature supporting the general validity of brief measures of these kinds of health risks (Bjorner et al., 2005; Cella et al., 2010; Kroenke & Spitzer, 2002; Magasi et al., 2011; Martin et al., 2006; Smith et al., 2009) and work factors (Jenkins, 2014; Kessler et al., 2003, 2004; Mitchell & Bates, 2011).

### **Implications for Practice**

Although successful, this project encountered some operational challenges. When concerns were identified, both the EAP staff and the on-site employees worked together to try to address them. Over the course of the clinical portion of the project (which lasted about two years), several operational concerns were successfully resolved, but one element remained an issue throughout the duration of the project. The task of introducing patients who screened positive to the counselor for a chance to begin intervention work (“introduction rate”) required ongoing adjustments. We came to realize that when patients were introduced to the counselor almost immediately after completing a screening, it often produced significant positive effects that impacted them in meaningful ways. When such an immediate introduction to the counselor was not possible, patients who screened positive may not have engaged in the counseling.

More generally, it is our view that the training (or more accurately – the lack of it) may be the single greatest barrier to providing interventions of this kind. In our experience, it takes a minimum of one full week of formal training coupled with a 6-month long period of supervision (including regular review and feedback of patient-counselor session recordings) before a provider can attain a sufficiently high level of proficiency in the BSI approach, which includes Motivational Interviewing and Collaborative Care. Such proficiency is needed to produce the outcomes reported herein. We contend that far too few practitioners are trained at this level. It is more common for behavioral health providers to attend one or two workshops in the techniques of Motivational Interviewing and then believe this to be sufficient preparation. We contend this is not enough training in this technique for it be delivered effectively. Further, additional training in the collaborative care models for the treatment of depression, anxiety and substance use that better align psychological and medical care resources (Community Preventive Services Task Force, 2012) is also of critical importance for BSI programs and this aspect too is often lacking for many medical care providers.

Lastly, although Screening, Brief Intervention and Referral to Treatment (SBIRT) programs have become more popular (McPherson et al., 2009), these tend to be limited by either a narrow focus on substance use issues or by not actually providing treatment interventions (they are often just focused on screening for risk and making a referral of those found to be at risk to other specialty provider programs). In contrast, we demonstrated in this project how screening for multiple risk factors, including depression, stress, smoking, physical exercise, nutrition, relationship, legal, financial, and other family problems can pay great dividends. If clinicians who perform these risk screenings also have adequate training in providing evidence-based treatment approaches (particularly the motivational interviewing approach), much can be

accomplished through brief interventions delivered before a referral is needed to a more expensive specialty provider of behavioral health services.

Given the demonstrated effectiveness of BSI, the intervention is one that should be actively considered by Accountable Care Organizations (ACOs), self-insured employers and pooled insurance trusts. ACOs, established under the Patient Protection and Affordable Care Act, are organizations of healthcare providers whose reimbursement is tied to quality outcomes measures and reduced health care costs. When ACOs maintain quality outcomes and reduce costs they are permitted to share those cost savings with the insurer, typically the Medicare program. An effective prevention program, such as BSI, would enable ACOs to do just that. Similarly, self-insured employers and pooled insurance trusts stand to save money when small expenditures on prevention programs, such as BSI, can produce enhanced health outcomes that could reduce longer-term medical expenditures. Again, BSI appears to be just such a program.

### **Conclusion**

We contend that this kind of BSI program appears promising, especially when involving onsite and convenient counseling delivered by properly trained clinicians. Results from our pilot study indicated that the risk finding process was effective in identifying adults appropriate for early intervention (i.e., over three-fourths of cases who focused on depression as an issue during treatment had initially screened above the clinical threshold for moderate severity) and also that health and work outcomes for participants were improved after counseling. The potential cost savings associated with the program in areas of health care costs and work performance could also be explored in future research.

---

### **Acknowledgments**

**Acknowledgments.** The authors want to recognize the support of Diane Dalmasse (Director of the State of Vermont Division of Vocational Rehabilitation), Joelle Judge, J.D. (Senior Project Manager on the Vermont Health Care Innovation Project), Georgia J. Maheras, Esq. (Deputy Director of Health Care Reform for Payment and Delivery System Reform and Director of the Vermont Health Care Innovation Project). The authors would like to thank our project manager Deborah Buley, Ph.D., and our two counselors Andy Kelley, M.S.W., and Susan Hall, M.A. (both from Invest EAP Centers for Wellbeing) for their contributions to the study. A special thank you to Richard Brown, M.D., M.P.H., University of Wisconsin, School of Medicine and Public Health, who provided important consultation and training. A preliminary report on this study was presented at the Vermont Worksite Wellness Conference (Attridge & Dickens, 2018).

**Funding.** Funding for this study was provided by the State of Vermont, Vermont Health Care Innovation Project, under Vermont's State Innovation Model (SIM) grant, awarded by the Center for Medicare and Medicaid Services (CMS) Innovation Center (CFDA Number 93.624) Federal Grant #1G1CMS331181-03-01.

**Disclaimer.** The contents provided are solely the responsibility of the authors and do not necessarily represent the official views of the State of Vermont or of the United States Department of Health and Human Services or any of its agencies.

**Conflict of Interest.** Dr. Attridge was hired by Invest EAP Centers for Wellbeing to conduct this research project. Steve Dickens is a senior executive at Invest EAP Centers for Wellbeing.

---



### References

- Agerwala, S. M., & McCance-Katz, E. F. (2012). Integrating screening, brief intervention, and referral to treatment (SBIRT) into clinical practice settings: A brief review. *Journal of Psychoactive Drugs, 44*(4), 307-317. <https://doi.org/10.1080/02791072.2012.720169>
- Attridge, M. (2012). Employee assistance programs: Evidence and current trends. In R. J. Gatchel & I. Z. Schultz (Eds.), *Handbook of occupational health and wellness* (pp. 441-467). New York, NY: Springer Press.
- Attridge, M. (2015, September 28-29). *EAP business value 2-day workshop: ROI paths, pricing and promises* [Professional training]. Employee Assistance Professionals Association annual conference, San Diego, CA, United States.
- Attridge, M. (2016, November 2). *EAP industry outcomes for employee absenteeism and presenteeism: A global research analysis*. Presented at the Employee Assistance Professionals Association annual conference, Chicago, IL, United States.  
<http://hdl.handle.net/10713/7203>
- Attridge, M. (2019). A global perspective on promoting workplace mental health and the role of employee assistance programs. *American Journal of Health Promotion, 34*(4), 622-627.  
<https://doi.org/10.1177/0890117119838101c>
- Attridge, M. & Dickens, S. (2018, March 21). *Health and emotional wellbeing in the workplace: Behavioral screening and intervention as an effective approach to wellness*. Presented at the Vermont Worksite Wellness Conference, Burlington, VT.  
<http://hdl.handle.net/10713/14857>
- Attridge, M., Rosenberg, T., Otis, J., Riedel, J., Lynch, W., & Sullivan, S. (2001, March 19). Presenteeism and health: Exploring the link between employee productivity and common

- physical and mental health issues. [White paper]. Scottsdale, AZ: Institute for Health and Productivity Management. <http://hdl.handle.net/10713/15067>
- Attridge, M., Sharar, D., DeLapp, G., & Veder, B. (2018). EAP Works: Global results from 24,363 counseling cases with pre-post data on the Workplace Outcome Suite. *International Journal of Health and Productivity*, 10(2), 5-25. <http://hdl.handle.net/10713/8962>
- Berndt, E. R., Finkelstein, S. N., Greenberg, P. E., Howland, R. H., Keith, A., Rush, A. J., ... & Keller, M. B. (1998). Workplace performance effects from chronic depression and its treatment. *Journal of Health Economics*, 17(5), 511-535. [https://doi.org/10.1016/S0167-6296\(97\)00043-X](https://doi.org/10.1016/S0167-6296(97)00043-X)
- Bjorner, J. B., Fayers, P. M., & Idler, E. L. (2005). Self-rated health. In P. M. Fayers & R. D. Hays (Eds.), *Assessing quality of life* (pp. 309–323). Oxford, England: Oxford University Press.
- Boles, M., Pelletier, B., & Lynch, W. (2004). The relationship between health risks and work productivity. *Journal of Occupational and Environmental Medicine*, 46, 737-745. [doi:10.1097/01.jom.0000131830.45744.97](https://doi.org/10.1097/01.jom.0000131830.45744.97)
- Bray, J. W., Zarkin, G. A., Davis, K. L., Mitra, D., Higgins-Biddle, J. C., & Babor, T. F. (2007). The effect of screening and brief intervention for risky drinking on health care utilization in managed care organizations. *Medical Care*, 45(2), 177–182. [doi:10.1097-01.mlr.0000252542.16255.fc](https://doi.org/10.1097-01.mlr.0000252542.16255.fc)
- Brown, R. (2016). *Personal discussion based on the experience of Wisconsin Initiative to Promote Healthy Lifestyles*. Madison, WI: University of Wisconsin.

- Brown, R., Moberg, D. P., Allen, J. B., Peterson, C. T., Saunders, L. A., Croyle, M. D., Lecoanet, R. M., Linna, S. M., Breidenbach, K., & Caldwell, S. (2014). Team approach to systematic behavioral screening and intervention. *American Journal of Managed Care*, 20(4), e113-e121. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4306562/>
- Bruffaerts, R., Vilagut, G., Demyttenaere, K., Alonso, J., AlHamzawai, A., Andrade, L.H., ...Kessler, R.C. (2012). Role of common mental and physical disorders in partial disability around the world. *British Journal Psychiatry*, 200(6), 454–461.  
doi:10.1192/bjp.bp.111.097519
- Cella, D., Riley, W., Stone, A., Rothrock, N., Reeve, B., Yount, S., Amtmann, D., Bode, R., Buysse, D., Choi, S., Cook, K., DeVellis, R., DeWalt, D., Fries, J. F., Gershon, R., Hahn, E. A., Jin-Shei Lai, J-S., ... Hays, R. D. on behalf of the PROMIS Cooperative Group. (2010). Initial item banks and first wave testing of the Patient–Reported Outcomes Measurement Information System (PROMIS) network: 2005–2008. *Journal of Clinical Epidemiology*, 63(11), 1179-1194. <https://doi.org/10.1016/j.jclinepi.2010.04.011>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. NJ: Lawrence Erlbaum Associates.
- Community Preventive Services Task Force. (2012). Recommendation from the Community Preventive Services Task Force for use of collaborative care for the management of depressive disorders. *American Journal of Preventive Medicine*, 42(5), 521-524.  
<https://doi.org/10.1016/j.amepre.2012.01.010>
- Employee Assistance Professionals Association. (2011). *Definitions of an employee assistance program and EAP core technology*. [White paper.]

<https://www.eapassn.org/About/About-Employee-Assistance/EAP-Definitions-and-Core-Technology>

- Goetzel, R. Z., Henke, R. M., Head, M. A., Benevent, R., & Rhee, K. (2020). Ten modifiable health risk factors and employees' medical costs: An update. *American Journal of Health Promotion, 34*(5), 490-499. <https://doi.org/10.1177/0890117120917850>
- Hays, R. D., Bjorner, J., Revicki, R. A., Spritzer, K. L., & Cella, D. (2009). Development of physical and mental health summary scores from the Patient Reported Outcomes Measurement Information System (PROMIS) global items. *Quality of Life Research, 18*(7), 873–880. <https://doi.org/10.1007/s11136-009-9496-9>
- Jenkins, K. R. (2014). How valid are self-reports of illness-related absence? Evidence from a university employee health management program. *Population Health Management, 17*(4), 211-217. doi:10.1089/pop.2013.0087
- Karg, R. S., Bose, J., Batts, K. R., Forman-Hoffman, V. L., Liao, D., Hirsch, E., Pemberton, M. R., Colpe, L. J., & Hedden, S. L. (2014, October). Past year mental disorders among adults in the United States: Results from the 2008–2012 Mental Health Surveillance Study. *CBHSQ Data Review*. Published by the Center for Behavioral Health Statistics and Quality, SAMHSA. <http://www.samhsa.gov/data/sites/default/files/NSDUH-DR-N2MentalDis-2014-1/Web/NSDUH-DR-N2MentalDis-2014.pdf>
- Kessler, R. C., Barber, C., Beck, A., Berglund, P., Cleary, P. D., McKeenas, D., Pronk, N., Simon, G., Stang, P., Üstün, T. U., Wang, P. (2003). The World Health Organization Health and Work Performance Questionnaire (HPQ). *Journal of Occupational and Environmental Medicine, 45*(2), 156-174. doi:10.1097/01.jom.0000052967.43131.51

- Kessler, R. C., Ames, M., Hymel, P. A., Loeppke, R., McKenas, Richling, D. E., Stang, P. E., & Ustun, T. B. (2004). Using the WHO Health and Work Performance Questionnaire (HPQ) to evaluate the indirect workplace costs of illness. *Journal of Occupational and Environmental Medicine*, 46(Suppl. 6), S23-S37.
- Kowlessar, N. M., Goetzel, R. Z., Carls, G. S., Tabrizi, M. J., & Guindon, A. (2011). The relationship between 11 health risks and medical and productivity costs for a large employer. *Journal of Occupational and Environmental Medicine*, 53(5), 468-477. doi:10.1097/JOM.0b013e31821586b8
- Kroenke, K. (2012). Enhancing the clinical utility of depression screening. *Canadian Medical Association Journal*, 184(3), 281-282. doi:10.1503/cmaj.112004
- Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: A new depression diagnostic and severity measure. *Psychiatric Annals*, 32(9), 509-515. <https://doi.org/10.3928/0048-5713-20020901-06>
- Lennox, R. D., Sharar, D., Schmitz, E., & Goehner, D. B. (2018). Validation of the 5-item short form version of the Workplace Outcome Suite©. *International Journal of Health and Productivity*, 10(2), 49-61. <http://hdl.handle.net/10713/8973>
- Lohaus, D., & Habermann, W. (2019). Presenteeism: A review and research directions. *Human Resource Management Review*, 29(1), 43-58. <https://doi.org/10.1016/j.hrmmr.2018.02.010>
- Madras, B. K., Compton, W. M., Avula, D., Stegbauer, T., Stein, J. B., & Clark, H. W. (2009). Screening, brief interventions, referral to treatment for illicit drug and alcohol use at multiple healthcare sites: Comparison at intake and 6 months. *Drug and Alcohol Dependence*, 99(1-3), 280-295. <https://doi.org/10.1016/j.drugalcdep.2008.08.003>

- Magasi, S., Ryan, G., Revicki, D., Lenderking, W., Hays, R. D., Brod, M., Synder, C., Bpers, M., & Cella, D. (2011). Content validity of patient outcomes: Perspectives from a PROMIS meeting. *Quality of Life Research, 21*(5), 739-746.  
<https://doi.org/10.1007/s11136-011-9990-8>
- Martin, A., Rief, W., Klaiberg, A., & Braehler, E. (2006). Validity of the brief patient health questionnaire mood scale (PHQ-9) in the general population. *General Hospital Psychiatry, 28*(1), 71-77. <https://doi.org/10.1016/j.genhosppsych.2005.07.003>
- McPherson, T. L., Goplerud, E., Olufokunbi-Sam, D., Jacobus-Kantor, L., Lusby-Treber, K. A., & Walsh, T. (2009). Workplace alcohol screening, brief intervention, and referral to treatment (SBIRT): A survey of employer and vendor practices. *Journal of Workplace Behavioral Health, 24*(3), 285-306. doi:10.1080/15555240903188372
- Milliman, Inc. (2020). *How do individuals with behavioral health conditions contribute to physical and total healthcare spending?* [White paper]. Authors: Davenport, S., Gray, T. J., & Melek, S. P. <https://www.milliman.com/-/media/milliman/pdfs/articles/milliman-high-cost-patient-study-2020.ashx>
- Mitchell, R. J., & Bates, P. (2011). Measuring health-related productivity loss. *Population Health Management, 14*(2), 93-98. <https://doi.org/10.1089/pop.2010.0014>
- Mitchell, R. J., Ozminkowski, R. J., & Serxner, S. (2013). Improving employee productivity through improved health. *Journal of Occupational and Environmental Medicine, 55*(10), 1142–1148. <http://doi.org/10.1097/JOM.0b013e3182a50037>
- Morneau Shepell. (2020). *Workplace Outcome Suite (WOS) annual report 2020: Part 1 - decade of data on EAP counseling reveals prominence of presenteeism*. Author: M. Attridge.

[White paper.] <https://wellbeing.lifeworks.com/wp-content/uploads/2020/09/WOS-annual-report-2020-part1-.pdf>

O'Donnell, M. P., Schultz, A. B., & Yes, L. (2015). The portion of health care costs associated with lifestyle-related modifiable health risks based on a sample of 223,461 employees in seven industries: The UM-HMRC Study. *Journal of Occupational and Environmental Medicine*, 57(12), 1284-1290. doi:10.1097/JOM.0000000000000600

Ornstein, S. M., Nietert, P. J., Jenkins, R. G., & Litvin, C. B. (2013). The prevalence of chronic diseases and multimorbidity in primary care practice: A PPRNet report. *Journal of the American Board of Family Medicine*, 26(5), 518–524. doi:0.3122/jabfm.2013.05.130012

PROMIS. (2017, March). *Global health: A brief guide to the PROMIS<sup>®</sup> global health instruments*.

[https://www.healthmeasures.net/images/PROMIS/manuals/PROMIS\\_Global\\_Scoring\\_Manual.pdf](https://www.healthmeasures.net/images/PROMIS/manuals/PROMIS_Global_Scoring_Manual.pdf)

Shi, Y., Sears, L. E., Coberley, C. R., & Pope, J. E. (2013). The association between modifiable well-being risks and productivity. *Journal of Occupational and Environmental Medicine*, 55(4), 353–364. doi:10.1097/JOM.0b013e3182851923

Smith, P. C., Schmidt, S. M., Allenworth-Davies, D., & Saitz, R. (2009). Primary care validation of a single-question alcohol screening test. *Journal of General Internal Medicine*, 24(7), 783-788. <http://doi.org/10.1007/s11606-009-0928-6>

Stewart, W. F., Ricci, J. A., Chee, E., & Morganstein, D. (2003). Lost productive work time costs from health conditions in the United States: Results from the American Productivity Audit. *Journal of Occupational and Environmental Medicine*, 45, 1234-1246.

- Thorpe, K., Jain, S., & Joski, P. (2017). Prevalence and spending associated with patients who have a behavioral health disorder and other conditions. *Health Affairs*, *36*(1), 124-132. doi:10.1377/hlthaff.2016.0875
- White, J., Hartley, S. K., Musich, S., Hawkins, K., & Ozminkowski, R. J. (2013). A more generalizable method to evaluate the association between commonly reported health risks and health care expenditures among employers of all sizes. *Journal of Occupational and Environmental Medicine*, *55*(10), 1179-1185. doi:10.1097/JOM.0b013e31829b2833
- White, K., Hedden, S., Palmer, J., & Arndt, S. (2014). Screening, brief intervention and referral to treatment (SBIRT): Year two annual evaluation report (Iowa Department of Public Health contract #5884YM50). Iowa Consortium for Substance Abuse Research and Evaluation. <http://iconsortium.subst-abuse.uiowa.edu/>
- Vermont Department of Health. (2017). *Vermont behavioral risk factor surveillance system: 2016 data summary*. Burlington, VT: Author. <https://www.healthvermont.gov/health-statistics-vital-records/population-health-surveys-data/brfss>
-



**Table 1.** Profile of clinical utilization factors and longitudinal data collection.

Measure	Time 1
Counseling provided onsite at clinic:	100%
Clinical issue (defined by client):	% ( <i>n</i> )
<b>Depression</b>	<b>57% (68)</b>
Smoking	22% (26)
Drinking alcohol	10% (12)
Diet and/or exercise	7% (8)
Other issue	5% (6)
Number of counseling sessions:	% ( <i>n</i> ) $M = 4.2$ (SD = 2.1)
1 session	9 (11)
2 sessions	12 (14)
3 sessions	23 (27)
<b>4 sessions</b>	<b>18 (22)</b>
5 sessions	17 (20)
6 sessions	9 (11)
7 sessions	4 (5)
8 sessions	3 (4)
9 sessions	3 (3)
10 sessions	3 (3)
Final scheduled session completed:	<b>Yes = 68%</b> ( <i>n</i> = 82); No = 32%
Session duration in minutes:	$M = 36.9$ (SD = 11.6)
Counselors:	<b>A (female) = 115 cases</b> ; B (male) = 5 cases
Time in months between data collection points:	
Time 1 to Time 2	$M$ (SD) = <b>2.8 months</b> (2.2) range: 1 to 11
Time 2 to Time 3 (target 3 months)	$M$ (SD) = <b>3.7 months</b> (0.9) range: 1 to 6
Time 2 to Time 4 (target 6 months)	$M$ (SD) = <b>6.8 months</b> (1.3) range: 3 to 10

*Note:* Total  $N = 120$ .  $M$  = mean; SD = standard deviation. Key characteristics in **bold**.

**Table 2.** Profile of client demographic and clinical utilization factors at baseline in full sample and at longitudinal time pairings

		Longitudinal Pairings				
		Time 1 Full Sample	Pair 1 Time 1 with Time 2	Pair 2 Time 1 with Time 3	Pair 3 Time 1 with Time 4	Average difference between pairings
<i>Client factors</i>	<i>n cases:</i>	120	84	85	49	
Age ( <i>M</i> years)		45.9	46.7	46.9	45.0	1.1 years
Race (% White)		94%	92%	94%	94%	1%
Gender (% females)		73%	75%	68%	76%	1%
Married (% Yes)		48%	52%	55%	57%	3%
Child at home (% Yes)		42%	42%	39%	33%	6%
Employed (% Yes)		38%	36%	40%	39%	2%
Household income ( <i>M</i> \$)		\$29,145	\$29,134	\$30,029	\$27,143	\$1,327
<i>Clinical factors</i>						
Sessions per case ( <i>M</i> )		4.2	4.8	4.3	4.7	0.7 sessions
Final session (% Yes)		68%	87%	67%	67%	13%
Counselor (% A)		96%	99%	98%	96%	2%
Clinical Issue						2%
Depression		57%	60%	60%	61%	
Smoking		22%	18%	16%	14%	
Drinking alcohol		10%	11%	11%	10%	
Diet & exercise		7%	8%	7%	8%	
Other		4%	3%	6%	7%	

*Note:* *M* = mean. Counselor A = female; Counselor B = male.

**Table 3.** Results for change after treatment on overall health outcome in full sample and employed subsample: By time pairing

		Time Point Longitudinal			
		1 Baseline	2 End case	3 Post 1 at 3 months	4 Post 2 at 6 months
<i>Full Sample</i>	<i>n cases:</i>	120	84	85	49
PROMIS-10	<i>M (SD)</i>	27.8 (6.9)	32.9 (6.6)	32.3 (5.7)	33.8 (7.8)
Tests of change after counseling		Pair 1: T1 27.4 (7.2) vs. T2 33.2 (6.3); paired $t(81) = -8.08, p < .001$ 21% improvement; Cohen $d = .89$ large effect			
		Pair 2: T1 27.9 (6.9) vs. T3 32.3 (5.8); paired $t(84) = -7.23, p < .001$ 16% improvement; Cohen $d = .79$ large effect			
		Pair 3: T1 26.7 (6.7) vs. T4 34.4 (7.3); paired $t(46) = -6.76, p < .001$ 29% improvement; Cohen $d = .99$ large effect			
<i>If Employed</i>	<i>n cases:</i>	46	26	32	17
PROMIS-10	<i>M (SD)</i>	29.4 (6.5)	35.4 (5.5)	31.0 (4.2)	33.3 (7.2)
Tests of change after counseling:		Pair 1: T1 29.1 (6.4) vs. T2 35.4 (5.5); paired $t(25) = -4.91, p < .001$ 22% improvement; Cohen $d = 0.96$ large effect			
		Pair 2: T1 30.3 (5.8) vs. T3 34.3 (4.5); paired $t(30) = -4.04, p < .001$ 13% improvement; Cohen $d = .71$ medium effect			
		Pair 3: T1 28.8 (7.0) vs. T4 36.9 (8.0); paired $t(16) = -3.42, p = .004$ 28% improvement; Cohen $d = .83$ large effect			

*Note:* Range 12 to 41; higher scores indicated better health. *M* = mean; *SD* = standard deviation.

**Table 4.** Results of tests of client background and clinical utilization factors as moderators of improvement in global health after treatment from baseline to 3-month follow-up

Factor ( <i>n</i> subgroups)	<i>F</i> value for factor X time interaction effect	% Improvement over baseline by subgroup
Overall Time Main Effect	52.36, $p < .001$	16% total sample average
<i>Zero-order Tests:</i>		
Client Background Factors:		
Race	NA	(sample sizes too imbalanced to test)
Age (years)	$< 1, p = .56$	NA
Gender (F=57, M=28)	2.99, $p = .09$	Female = 19%; Male 10%
Married (Yes=46, No=39)	2.29, $p = .13$	Married = 19%; Not = 13%
Child at home (Yes=33, No=52)	2.25, $p = .14$	Child home = 14%; Not = 18%
Employed (Yes=34, No=51)	$< 1, p = .76$	Employed = 14%; Unemployed = 18%
Household income (\$) <sup>a</sup>	7.93, $p = .01$	<b>Low = 24%; Med. = 17%; High = 1%</b>
Clinical Factors:		
Counselor A or B	NA	(sample sizes too imbalanced to test)
Counseling sessions (1-10) <sup>b</sup>	5.43, $p = .02$	<b>Low = 9%; Med. = 16%; High = 21%</b>
Final session (Yes=58, No=21)	$< 1, p = .65$	Clinical closure = 15%; Not = 18%
Clinical issue: (5 types)	1.75, $p = .15$	Depression ( $n = 51$ ) = 17% Smoking ( $n = 14$ ) = 7% Drinking alcohol ( $n = 9$ ) = 8% Diet & exercise ( $n = 6$ ) = 13% Other ( $n = 5$ ) = 32%
<hr/>		
<i>Multivariate Test as Set</i>		<i>Adjusted for Other Two Moderator Factors:</i>
Household income	7.18, $p = .01$	<b>Low = 20%; Med. = 16%; High = 1%</b>
Between-Subjects Effect	7.62, $p = .01$	<i>M</i> Pre: Low = 27%; Med = 27%; <b>High = 35%</b> <i>M</i> Post: Low = 32%; Med = 32%; <b>High = 36%</b>
Counseling sessions	4.99, $p = .03$	<b>Low = 8%; Med. = 17%; High = 20%</b>
Between-Subjects Effect	$< 1, p = .55$	
Gender	2.11, $p = .15$	Female = 18%; Male 11%
Between-Subjects Effect	$< 1, p = .66$	

*Note:*  $N = 85$ . Main effect  $F$  value for Pre Time 1 vs. Post Time 3 change and percentage improvement shown on line 1 is without any moderators in the model. Other  $F$  values for interaction effect of Time X moderator factor (d.f. = 1, 84).  $F$  values for between-subjects group effect (d.f. = 1, 81). PROMIS-10 score range 10-50.

<sup>a</sup> Income tested as variable with full five groups, but shown above as three groups: Low = \$15,000 or less ( $n = 53$ ); Medium = \$30,000 or \$42,500 ( $n = 22$ ); High = \$62,500 or \$85,000+ ( $n = 10$ ).

<sup>b</sup> Number of counseling sessions used tested as full range of 1 to 10, but shown above as three groups: Low = 1-2 sessions ( $n = 16$ ); Medium = 3-5 sessions ( $n = 48$ ); High = 6 to 10 sessions ( $n = 21$ ).

**Table 5.** Results for change after treatment on depression outcome among subsample of cases with clinical issue of depression: By time pairing

		Time Point Longitudinal			
		1 Baseline	2 End case	3 Post 1 at 3 months	4 Post 2 at 6 months
<i>If Clinical Issue of Depression</i>	<i>n cases:</i>	68	50	52	29
PHQ-9 > 10	<i>At-risk %</i>	74%	22%	35%	14%
PHQ-9 score	<i>M (SD)</i>	14.3 (6.1)	7.2 (6.4)	7.5 (5.4)	5.4 (6.4)

Pair 1: T1 14.9 (6.2) > T2 7.5 (6.5); paired  $t(45) = 8.27, p < .001$

50% improvement; Cohen  $d = 1.22$  large effect

Tests of change  
after counseling  
in PHQ-9 scores:

Pair 2: T1 14.1 (6.2) > T3 7.8 (5.5); paired  $t(48) = 7.43, p < .001$

45% improvement; Cohen  $d = 1.06$  large effect

Pair 3: T1 15.2 (6.0) > T4 5.8 (6.6); paired  $t(25) = 9.44, p < .001$

62% improvement; Cohen  $d = 1.85$  large effect

*Note:* PHQ-9 = Patient Health Questionnaire 9-item measure of depression. Range from 0 to 27; higher scores indicated greater depression symptoms. Scores of 10 or higher indicate clinical risk level (moderate or higher severity). *M* = mean; *SD* = standard deviation.

**Table 6.** Results for change after counseling in work presenteeism among employed subsample: By time pairing

		Time Point Longitudinal			
		1 Baseline	2 End case	3 Post 1 at 3 months	4 Post 2 at 6 months
<i>If Employed</i>	<i>n</i> cases:	46	26	32	17
Presenteeism Problem	% cases	33%	15%	3%	6%
Presenteeism Score	<i>M</i> (SD)	3.1 (1.0)	2.4 (3.1)	2.1 (1.0)	2.0 (0.8)

Change after counseling:

Pair 1: T1 3.1 (0.9) vs. T2 2.4 (1.0); paired  $t(25) = 3.14, p = .004$   
 23% improvement; Cohen  $d = .62$  medium effect

Pair 2: T1 3.0 (0.9) vs. T3 2.1 (1.0); paired  $t(31) = 4.00, p < .001$   
 30% improvement; Cohen  $d = .71$  medium effect

Pair 3: T1 3.5 (0.6) vs. T4 2.0 (0.8); paired  $t(16) = 6.93, p < .001$   
 43% improvement; Cohen  $d = 1.68$  large effect

*Note:* Rated from 1 to 5; higher score indicates greater presenteeism (i.e., not focusing fully on work). *M* = mean; *SD* = standard deviation. Problem for presenteeism if experienced “most” or “all of the time” in past four weeks (rating of 4 or 5).

**Table 7.** Results for change after treatment in hours of lost productive work time per month among employed subsample: By longitudinal time pairing

		Time Point Longitudinal			
		1 Baseline	2 End case	3 Post 1 at 3 months	4 Post 2 at 6 months
<i>If Employed</i>	<i>n cases:</i>	46	26	32	17
Hours of work absenteeism <sup>a</sup>		11.2 (22.4)	10.5 (25.5)	6.9 (12.0)	0.6 (1.8)
Hours worked from 160 normal monthly schedule		148.8	149.5	153.1	159.4
Productivity level while working (0-10) <sup>ab</sup>		5.4 (2.1)	7.2 (2.0)	7.2 (2.1)	7.5 (1.8)
% time unproductive while working		46%	28%	28%	25%
Hours of time unproductive while working		65.5	39.8	41.9	39.2
Combined hours of lost productive time (LPT)		76.7 (35.5)	50.3 (39.0)	48.8 (36.2)	39.8 (30.5)

Pair 1: T1 78.0 (38.9) vs. T2 50.3 (39.0); paired  $t(25) = 4.13, p < .001$

36% improvement; Cohen  $d = .81$  large effect

Change over time  
in LPT hours:

Pair 2: T1 73.1 (36.5) vs. T3 48.8 (36.2); paired  $t(31) = 3.24, p = .003$

33% improvement; Cohen  $d = .57$  medium effect

Pair 3: T1 74.5 (31.7) vs. T4 39.8 (30.5); paired  $t(16) = 4.08, p = .001$

47% improvement; Cohen  $d = .99$  large effect

*Note:*  $M$  = mean;  $SD$  = standard deviation.

<sup>a</sup> self-report item on survey; other measures calculated by authors.

<sup>b</sup> Rated from 0 (worst) to 10 (top performance).



**Table 8.** Results for associations between improvement after counseling in overall health and improvement in hours of lost productive work time per month among employed subsample: By time pairing

		Longitudinal Pairing		
		Pair 1 Time 1 with Time 2	Pair 2 Time 1 with Time 3	Pair 3 Time 1 with Time 4
<i>If Employed</i>	<i>n cases:</i>	26	32	17
Average improvement from baseline to later time (as percentage):				
	Overall health	22%	13%	28%
	Work LPT hours	36%	43%	47%
Correlation of improvement from baseline to later time on overall health with improvement from baseline to later time on LPT hours		$r = .33$ $p = .10$	$r = .35$ $p = .05$	$r = .50$ $p = .04$

*Note:* Correlation of standardized z-scores of percentage change over time after treatment.