Effect of Motivational Interviewing-Based Health Coaching on Employees’ Physical and Mental Health Status

Susan Butterworth, Ariel Linden, Wende McClay, and Michael C. Leo
Oregon Health & Science University

Motivational Interviewing (MI) based health coaching is a relatively new behavioral intervention that has gained popularity in public health because of its ability to address multiple behaviors, health risks, and illness self-management. In this study, 276 employees at a medical center self-selected to participate in either a 3-month health coaching intervention or control group. The treatment group showed significant improvement in both SF-12 physical ($p = .035$) and mental ($p < .0001$) health status compared to controls. Because of concerns of selection bias, a matched case-control analysis was also performed, eliciting similar results. These findings suggest that MI-based health coaching is effective in improving both physical and mental health status in an occupational setting.

Keywords: motivational interviewing, worksite wellness, health status

The precipitous rise in health care costs has strained the resources available to finance the system, including private employer-sponsored health insurance coverage (The Henry J. Kaiser Family Foundation, 2004). In 2004, companies reported increases of 12% on insurance premiums, and 24% of employers reported that they planned to raise employee contributions, while 30% planned to raise dependent contributions (Hewitt, 2004). This increased financial burden on employees is significant as their share of the health care dollar has increased 126% over the last five years, compared to 76% for employers (Hewitt, 2004). Even worse, some employers are eliminating health care coverage for their employees altogether. In fact, more than a quarter of all firms with more than 500 employees do not offer any employer-based health insurance for workers and their families (Collins, Davis, Doty, & Ho, 2004).

Behavior-related health practices are linked, either directly or indirectly, to these health care costs, in addition to absenteeism, presenteeism, and productivity (Goetzel et al., 2004, 1998; Haynes & Dunagan, 2002). Moreover, behavioral choices are among the top health indicators for morbidity and mortality in the United States, such as physical activity, poor dietary patterns, tobacco use, and substance abuse. (U.S. Department, 2000). Mental health conditions from anxiety, stress, and depression are significant cost burdens in the workplace, are a leading cause of disability, and have historically been underreported and treated (Goetzel, Hawkins, Ozminkowski, & Wang, 2003; Kalia, 2002; National Institute of Occupational Safety & Health [NIOSH], 2004; Stewart, Ricci, Chee, Hahn, & Morganstein, 2003). Stress is cited as being responsible for 12% of all unscheduled absences in the workplace (CCH, 2005).

As employers are now well aware of this link (Goetzel et al., 2004), behavior management through worksite wellness programs has become a popular and effective means to reduce health risks and employer costs, increase productivity, and improve quality of life (Aldana & Pronk, 2001; Goetzel, Juday, & Ozminkowski, 1999; Keyes & Grzywacz, 2005). Typical interventions include health risk assessments, educational programs, biometric screenings (e.g., measurements of blood pressure, blood lipids, BMI, etc.) and health coaching.

Health coaching is a relatively new behavioral intervention that has gained popularity in health promotion, public health, and disease management because of the ability to address multiple behaviors, health risks, and self-management of illness in a cost-effective manner. In the context of the current study, health coaching is defined as a service in which providers facilitate participants in changing lifestyle-related behaviors for improved health and quality of life, or establishing and attaining health-promoting goals (Van Ryn & Heaney, 1997). Moti-
vational Interviewing (MI) is an evidence-based approach that is being increasingly incorporated in the health coaching process (Bennett et al., 2005; Hecht et al., 2005; Linden, Butterworth, & Roberts, 2006; Miller, 2004).

MI was originally developed for addictions counseling in the 1980s and is described as a “directive, client-centered counseling style for eliciting behavior change by helping clients to explore and resolve ambivalence” (Miller, 1983; Miller & Rollnick, 2002). As demonstrated in two recent meta-analyses, MI has been shown to be effective for treating addictions such as illegal drugs, smoking, and alcoholism (Hettema et al., 2005; Miller & Rollnick, 2002). The practitioner or coach emphasizes the three underlying assumptions of MI—collaboration, evocation, and autonomy—in order to establish rapport, reduce resistance, and elicit “change talk” (one’s own reasons and arguments for change) (Hettema et al., 2005; Miller & Rollnick, 2002). The intended outcome of these MI sessions is for clients to resolve ambivalence (a central goal), move through the stages of change (Prochaska, 1979), and follow through on desirable lifestyle change, which would ideally result in improved health outcomes.

In this study, we evaluated the impact of MI-based health coaching on the physical and mental health status of employees at a large medical university in the Northwest. This study was the precursor to a large randomized controlled trial that has recently been initiated. We hypothesized that survey data would show improvement in both mental and physical health status of employees who participated in health coaching over a 3-month period compared with those who did not.

Methods

Setting

The present study was conducted at Oregon Health and Science University (OHSU), a large health and research university in the Pacific Northwest that employs over 11,000 workers. In 1998, OHSU established an Employee Wellness Program (EWP), focused on evidence-based prevention methods. The program is made available to over 9000 employees as part of their benefits package and includes targeted communications, health risk assessments, biometric screenings, support groups, health coaching, and other related offerings. From previous health risk assessment, pharmacy, and disability data gathered at this institution, it has been established that employees experience high stress levels and worse than average mental health. This is reflective of the unfavorable work conditions found in similar occupational settings (DiGiacomo & Adamson, 2001; Felton, 1998).

Participants

Randomization was not performed in this study as the objective was to incorporate the intervention into the existing EWP. OHSU employees were recruited into the study via web site announcements, posters, or verbal referral. Employees were also recruited into the control group by active enrollment methods in high traffic areas with the incentive of a Power Bar®. Participation required individuals to be current OHSU employees working 20 or more hours per week. Employees were excluded if they were either not eligible for benefits or had received health coaching within the last year from the EWP. One hundred forty-five participants self-selected into the treatment group, and 131 participants elected to participate in the control group.

Intervention

Treatment group participants were given a 3-month health coaching intervention with a minimum of one initial session and two follow-up contacts. Participants themselves determined the actual number of sessions they received based on need and interest. Health coaching was conducted by health care professionals rigorously trained in MI and
evaluated for proficiency by an independent coder with expertise in the Motivational Interviewing Skill Code (MISC) tool (Moyers, Martin, Catley, Harris, & Ahluwalia, 2003). Each session was limited to 30 minutes in duration. Presenting issues were recorded as primary and secondary health concerns and focused on typical public health issues such as weight loss, fitness, stress, and nutrition. However, as has been shown in the literature, coaching topics tend to be fluid and overlapping (Bennett et al., 2005); for example, stress-related issues consistently ran parallel with other presenting health issues. Control group participants received no intervention during the 3-month period but were offered health coaching at the end of the study.

**Outcome Measures**

The Short Form 12 version 2 (SF-12®) Health Survey (Ware et al., 1996) was administered to all study participants at commencement and the end of each subject’s 3-month study period. For control group participants at baseline, the survey was either administered in person or by mail, whereas postdata collection was completed by mail. For treatment group participants, the survey was administered in person at baseline and either in person or by mail poststudy. Because participation in either arm of the study was contingent upon completing the survey, the response rate was 100% at baseline. Poststudy response rate was 83.4% for treatment and 90% for control. The low loss to attrition is attributable to the short nature of the study, the high satisfaction level of treatment participants, and the brevity of the survey.

This health status survey is commonly used, brief (12 questions), and provides a description of the respondent’s health. This survey is based upon the SF-36 Health Survey (Ware & Sherbourne, 1992) and has at least one question from each of the SF-36’s original eight domains: physical functioning, role limitations due to physical functioning, general health perception, bodily pain, social functioning, energy/vitality, role limitations due to emotional functioning, and mental health. Two composite scores are derived from aggregating the responses to the 12 survey questions: the Mental Composite Score (MCS), and the Physical Composite Score (PCS). These are standardized to a 0 to 100 point scale to allow for comparisons across various populations with a general standard deviation of 10 (Ware, Kosinski, Turner-Bowker, & Gandek, 2002).

The primary analysis compared PCS and MCS scores between treatment and control groups. Given that the study was not randomized, selection bias was of major concern. To mitigate this threat to validity, a retrospective case-control design was also implemented in which a subset of 44 treatment and control group participants were matched on baseline characteristics using the propensity scoring technique (Linden, Adams, & Roberts, 2005). Although this method allows the researcher to control for known variation between matched pairs, unknown sources of bias may still exist. To address this concern, a sensitivity analysis (Linden, Adams, & Roberts, 2006) was further conducted to assess the amount of bias that would have to be present in order to nullify any results indicative of program effectiveness.

Between-groups comparisons were performed using two-sample *t* tests assuming equal variances. Within-group post analyses were conducted using paired *t* tests. Paired *t* tests were also used in all matched case-control analyses. An intraclass correlation was used to compare the distribution of subject-level characteristics between groups. Ordinary least squares regression was used to identify subject-level characteristics associated with a change in PCS and MCS scores. Statistical analyses were conducted using StatsDirect statistical software (www.statsdirect.com), and the level of significance was set a priori at *p* ≤ .05.

**Results**

Characteristics of study participants are shown in Table 1. Of the 276 participants initially enrolled, 37 were lost to attrition (24 and 13, for treatment and control groups respectively) resulting in a nearly even number of participants in each group. Upon analysis, those participants leaving the study early did not appear to differ in characteristics from those remaining in their respective cohorts.

There were no significant differences between the groups in the average age, the average job tenure, or distribution of participants by job category. However, there were significantly fewer males in the treatment group compared to controls (*p* < .0001), and, more importantly, the treatment group scored significantly lower on both the PCS and the MCS than the control group on the preprogram SF-12 (*p* = .001 and *p* < .0001 for PCS and MCS respectively).

The treatment group improved their outcomes on both the PCS (1.69 points, *p* = .035) and on the MCS (4.40 points, *p* < .0001), while the control group showed no statistically significant change on either scale. Inasmuch as these results appear to provide compelling support for the wellness program’s effectiveness, the differences between groups on baseline characteristics and SF-12 scores suggest that selection bias may have threatened the validity of those outcomes.

In an effort to control for bias, a case-control design was implemented in which treatment group participants were matched with controls on the propensity score. Table 2 shows the baseline characteristics that were used to create the propensity scores on which the 44 pairs were matched. As expected, the groups were well matched as indicated by no significant differences being noted on any baseline characteristic, most importantly the PCS and MCS.

The 44 cases had a similar increase on the PCS as did the treatment group of 121 (1.58 vs. 1.69); however, the sample size was not sufficiently large enough for this increase to prove statistically significant. On the other hand, cases increased their MCS scores by 3.45 points, which was a sufficiently large
effect size to demonstrate statistical significance (p = 0.016). Consistent with findings in the first analysis, the matched-controls did not show any statistically significant change in either their PCS or MCS scores. As the case-control method matched participants only on known baseline characteristics, there was still a concern that the MCS scores attained could be influenced by unknown bias. A sensitivity analysis

Table 1

<table>
<thead>
<tr>
<th>Variable</th>
<th>Treatment</th>
<th>Control</th>
<th>p-Value</th>
<th>95% CI for mean difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enrolled N</td>
<td>145</td>
<td>131</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lost to attrition</td>
<td>24</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ending N</td>
<td>121</td>
<td>118</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (yrs)</td>
<td>40.10</td>
<td>39.80</td>
<td>0.810</td>
<td>(−3.100, 2.450)</td>
</tr>
<tr>
<td>Male (%)</td>
<td>9.90</td>
<td>37.30</td>
<td>&lt;0.0001</td>
<td>(0.170, 0.378)</td>
</tr>
<tr>
<td>Tenure (yrs)</td>
<td>5.86</td>
<td>7.07</td>
<td>0.207</td>
<td>(−3.092, 0.673)</td>
</tr>
<tr>
<td>Job category (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Faculty/academic/research</td>
<td>25.62</td>
<td>40.68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Administrative classified</td>
<td>26.45</td>
<td>16.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Administrative unclassified</td>
<td>21.49</td>
<td>12.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physician</td>
<td>0.83</td>
<td>1.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nurse</td>
<td>5.79</td>
<td>5.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computer support</td>
<td>5.79</td>
<td>6.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facilities support</td>
<td>2.48</td>
<td>4.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>11.57</td>
<td>12.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre PCS</td>
<td>49.26</td>
<td>53.12</td>
<td>0.001</td>
<td>(−6.110, −1.620)</td>
</tr>
<tr>
<td>Pre MCS</td>
<td>43.30</td>
<td>49.49</td>
<td>&lt;0.0001</td>
<td>(−8.840, −3.540)</td>
</tr>
<tr>
<td>Post PCS</td>
<td>50.95</td>
<td>53.78</td>
<td>0.009</td>
<td>(−4.950, −0.700)</td>
</tr>
<tr>
<td>Post MCS</td>
<td>47.70</td>
<td>49.06</td>
<td>0.266</td>
<td>(−3.790, 1.050)</td>
</tr>
</tbody>
</table>

Note. CI = confidence interval; PCS = physical composite score; MCS = mental composite score.

a Intra-class correlation coefficient. b 95% Limits of agreement.

Table 2

<table>
<thead>
<tr>
<th>Variable</th>
<th>Cases (N = 44)</th>
<th>Controls (N = 44)</th>
<th>p-Value</th>
<th>95% CI for mean difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Propensity score</td>
<td>0.51</td>
<td>0.51</td>
<td>0.16</td>
<td>(−0.002, 0.002)</td>
</tr>
<tr>
<td>Age (yrs)</td>
<td>39.84</td>
<td>37.68</td>
<td>0.36</td>
<td>(−6.850, 2.530)</td>
</tr>
<tr>
<td>Male (%)</td>
<td>0.25</td>
<td>0.20</td>
<td>0.42</td>
<td>(−0.158, 0.067)</td>
</tr>
<tr>
<td>Tenure (yrs)</td>
<td>6.53</td>
<td>5.80</td>
<td>0.61</td>
<td>(−3.580, 2.130)</td>
</tr>
<tr>
<td>Job category (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Faculty/academic/research</td>
<td>16.95</td>
<td>11.57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Administrative classified</td>
<td>6.78</td>
<td>8.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Administrative unclassified</td>
<td>3.39</td>
<td>5.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physician</td>
<td>0.00</td>
<td>0.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nurse</td>
<td>0.85</td>
<td>1.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computer support</td>
<td>2.54</td>
<td>2.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facilities support</td>
<td>0.00</td>
<td>2.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>5.93</td>
<td>3.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre PCS</td>
<td>49.97</td>
<td>51.51</td>
<td>0.41</td>
<td>(−2.150, 5.230)</td>
</tr>
<tr>
<td>Pre MCS</td>
<td>48.08</td>
<td>47.88</td>
<td>0.92</td>
<td>(−4.230, 3.830)</td>
</tr>
<tr>
<td>Post PCS</td>
<td>51.55</td>
<td>52.52</td>
<td>0.58</td>
<td>(−2.520, 4.460)</td>
</tr>
<tr>
<td>Post MCS</td>
<td>51.53</td>
<td>46.83</td>
<td>0.03</td>
<td>(−8.840, −0.560)</td>
</tr>
</tbody>
</table>

Note. CI = confidence interval; PCS = physical composite score; MCS = mental composite score.

a Intra-class correlation coefficient. b 95% limits of agreement.
was conducted on the MCS scores to provide an estimate of how far this bias must diverge from the 50/50 split of a randomized controlled trial to raise concerns about the validity of the study findings. Following the procedures described by Linden, Adams, and Roberts (2006), the results of the sensitivity analysis suggest that intervention participants would need to be 1.28 times more likely to possess hidden bias than their matched controls in order to change our conclusion that the intervention led to significantly increased MCS scores. This value suggests that the current study results are relatively insensitive to the amount of bias necessary to alter our conclusions that the improvement in MCS scores was indeed an outcome of the program intervention and not a function of hidden bias.

Regression analysis failed to find an association between the independent variables (age, duration of tenure, job category, gender, number of health coaching sessions) and the two dependent variables (change from pre to postMCS and change from pre to postPCS scores) for the 121 participants in the intervention.

Discussion

One past criticism of worksite wellness programs has been that, without significant incentives, employees with lower health risk tend to participate more than employees at higher risk (Serxner, Anderson, & Gold, 2004). In contrast, this project demonstrated that the employees who self-selected into the intervention group were at higher risk than those self-selecting into the control group (i.e., they had significantly lower mental health status and function scores at baseline). This finding is what employers would hope for when implementing a health promotion intervention. Thus, one important tenet of this study is that it was implemented in a real world setting, and employees at greatest risk sought help without incentives. Recruitment and marketing efforts made an appeal to those who were interested in achieving health goals, and the phrase “health coaching” was used exclusively, versus “health counseling.” Several employees in the treatment group also remarked that they participated in order to “help us with our research study.”

This health promotion study also proved to be cost-effective. Although participants determined for themselves how many coaching sessions they would receive, it is worthy of note that the number of health coaching sessions did not independently influence the results, and the average number of sessions was only 2.7 (SE = 0.16). In addition, although mental health was significantly improved, the majority of the health coaching staff was comprised of health promotion specialists with training in behavior change and MI, as opposed to more costly mental health counselors or nurse specialists.

Although MI-based interventions have been shown to improve clinical outcomes in various settings and conditions (Rubak, Sandbæk, Lauritzen, & Christensen, 2005), this is the first published study in which health status (as measured by the SF-12) was studied as an outcome. The significance of this is that health status has been demonstrated to correlate well with medical expenditures and use of health care services (Fleishman, Cohen, Manning, & Kosinski, 2006). Similarly, low mental health composite scores show close correlation with clinical depression (Noel et al., 2004) and are indicators of stress and anxiety, which have been demonstrated as playing a crucial role in chronic disease (Chandola, Brunner, & Marmot, 2006) and workplace injuries (Swaen, van Amelsvoort, Bultmann, Slangen, & Kant, 2004). Therefore, in the absence of other clinical or cost indices, health status may serve well as a proxy.

Other characteristics of the MI technique which make it particularly suitable for use in a worksite behavioral change program are: (1) it is most effective when implemented with clients who are considered difficult; that is, reluctant to change, stuck, or ambivalent about changing their behavior; (2) it has been found to be efficacious in small doses; (3) it has been found to work across gender, age, cultural, and socioeconomic boundaries; and (4) it works well in conjunction with other traditional programs and interventions (Hettema et al., 2005).

An important issue for future studies to address is the mechanism by which MI-based health coaching impacts health status/function, particularly in the area of mental health. The average MCS for the treatment group at baseline was 43.30. To put this into perspective, the average MCS for the general U.S. population is 49.37. In fact, this value is lower than the norms for every chronic condition other than clinical depression, as surveyed in 1998 and reported by Ware et al. (2002). On average, the treatment group increased their MCS score by 4.4 points, which is the equivalent to almost half a standard deviation. In contrast, the average PCS for the treatment group at baseline (49.97), while still significantly lower than the baseline control Group PCS, compares favorably with the norm of the general U.S. population (Ware et al., 2002). Thus, even though average PCS did improve
in the treatment group, there was not as much room for improvement as there was in their MCS.

It is becoming more widely acknowledged that most lifestyle changes are infused with psychosocial dynamics such as ambivalence, self-efficacy, self-image, motivation, self-doubt, and core identity (Bandura, 2004; Bodenheimer, Lorig, Holman, & Grumback, 2005; Holahan & Suzuki, 2004; Loeb, 2004; Miller & Rollnick, 2002). As described by Prescott:

…MI views people as complex, driven by competing motives and in conflict with themselves. This complexity is noticeable in motivational conflict (ambivalence) and fluctuating levels of self-efficacy (both optimism and doubts about being able to change grow and fade). (2006, p. 7)

Thus it appears that MI is particularly well-suited for impacting the psychosocial aspects of behavior change.

Although every effort was taken to control for bias, limitations remain due to the quasi-experimental design that was used in this study. Selection bias was apparent as indicated by the differences noted between treatment and control groups on their baseline characteristics. Given this, we conducted a matched case-control analysis where participants were statistically matched on the propensity score. This method increases our confidence in the results; however, unknown sources of bias may still remain. The sensitivity analysis further supports that the results were not sensitive to bias. Nonetheless, without randomization no amount of statistical control will ensure that bias did not influence the treatment group differentially than controls. The small sample size in the case-control design reduced the statistical power to detect the same statistically significant improvement in PCS scores that was found in the primary analysis. Future studies implementing the matched case-control design should have sufficient group sizes to reduce the possibility of committing a Type II error. Finally, this study may not be generalizable across people, settings, treatments, or outcomes (Linden, Adams, & Roberts, 2004). This worksite has unique characteristics in that it is a medical university in the northwest with a low percentage of minorities, 60% women, and, although hosting a diverse population, a higher than average knowledge about health issues, and lower than average mental health scores.

Conclusion

The purpose of this study was to evaluate the impact of MI-based health coaching on the physical and mental health status of employees at a large worksite. It is of significance that the study was set in a real world setting where employees who self-selected into the health coaching intervention were nonincentivized but still at high risk. This is the first known published study in which health status and function (as measured by the SF-12 Health Survey) was used as an outcome for an MI-based intervention. Currently, a large-scale randomized controlled trial is being implemented at this same worksite as a follow up to this study. If the results can be replicated, this would further support MI-based health coaching as an effective health promotion intervention, and health status could serve well as a proxy in the absence of other clinical or cost indices. Perhaps of greatest interest is the mechanism by which MI influences mental health status and should be the emphasis in future studies.

References


