Combinations of Techniques That Effectively Change Health Behavior: Evidence From Meta-CART Analysis

Elise Dusseldorp, Netherlands Organization for Applied Scientific Research (TNO), Leiden, the Netherlands, and Katholieke Universiteit Leuven

Stef van Buuren, Netherlands Organization for Applied Scientific Research (TNO), Leiden, the Netherlands, and Utrecht University

Marieke W. Verheijden and Pepijn van Empelen, Netherlands Organization for Applied Scientific Research (TNO), Leiden, the Netherlands

Objective: Many health-promoting interventions combine multiple behavior change techniques (BCTs) to maximize effectiveness. Although, in theory, BCTs can amplify each other, the available meta-analyses have not been able to identify specific combinations of techniques that provide synergistic effects. This study overcomes some of the shortcomings in the current methodology by applying classification and regression trees (CART) to meta-analytic data in a special way, referred to as Meta-CART. The aim was to identify particular combinations of BCTs that explain intervention success.

Method: A reanalysis of data from Michie, Abraham, Whittington, McAteer, and Gupta (2009) was performed. These data included effect sizes from 122 interventions targeted at physical activity and healthy eating, and the coding of the interventions into 26 BCTs. A CART analysis was performed using the BCTs as predictors and treatment success (i.e., effect size) as outcome. A subgroup meta-analysis using a mixed effects model was performed to compare the treatment effect in the subgroups found by CART. Results: Meta-CART identified the following most effective combinations: Provide information about behavior–health link with Prompt intention formation (mean effect size $g = 0.46$), and Provide information about behavior–health link with Provide information on consequences and Use of follow-up prompts ($g = 0.44$). Least effective interventions were those using Provide feedback on performance without using Provide instruction ($g = 0.05$). Conclusions: Specific combinations of BCTs increase the likelihood of achieving change in health behavior, whereas other combinations decrease this likelihood. Meta-CART successfully identified these combinations and thus provides a viable methodology in the context of meta-analysis.

Keywords: intervention effectiveness, behavior change techniques, synergistic effects, classification and regression trees, subgroup, meta-analysis

Interventions to change health-related behaviors often include multiple behavior change techniques (BCTs) that are assumed to interact or have a cumulative effect, with the aim of maximizing the effectiveness of an intervention (Craig et al., 2008; Malotte et al., 2000). Thus far, heterogeneity in the effectiveness of health-related interventions is observed, and it is expected that differences in the techniques used by interventions may account for this heterogeneity. BCTs, such as Prompt intention formation and Provide feedback on performance, can be considered as the atomic parts of an intervention. Interventions may differ greatly in number and type of BCTs. A good theoretical understanding is needed concerning when and how interventions cause changes in health-related behaviors. Further research is needed to determine whether specific combinations of BCTs are effective for particular interventions.

The research reported in this paper was supported by a grant from “Platform Sport, Physical Activity and Education” [Platform Sport, Bewegen en Onderwijs], project Effectiveness of Interventions, the Netherlands, and supported by the Fund for Scientific Research – Flanders (project G.0546.09), Belgium. The authors gratefully acknowledge Susan Michie, Charles Abraham, Craig Whittington, and colleagues for making available their meta-analytic data for reanalysis, and Theo Paulussen, Paula van Dommelen, and two anonymous reviewers for helpful comments on a previous version of this article.

Correspondence concerning this article should be addressed to Elise Dusseldorp, TNO: Expertise Group Life Style, P. O. Box 2215, 2301 CE Leiden, the Netherlands. E-mail: elise.dusseldorp@tno.nl

This article was published Online First November 25, 2013.

Elise Dusseldorp, Expertise Group Life Style, Netherlands Organization for Applied Scientific Research (TNO), Leiden, the Netherlands, and Department of Psychology, Katholieke Universiteit Leuven, Leuven, Belgium; Lenneke van Genugten, Expertise Group Life Style, Netherlands Organization for Applied Scientific Research (TNO), and Department of Public Health, Erasmus, University Medical Centre Rotterdam, the Netherlands; Stef van Buuren, Expertise Group Life Style, Netherlands Organization for Applied Scientific Research (TNO), and Department of Methodology and Statistics, Faculty of Social and Behavioral Sciences, Utrecht University, Utrecht, the Netherlands; Marieke W. Verheijden and Pepijn van Empelen, Expertise Group Life Style, Netherlands Organization for Applied Scientific Research (TNO).
related behavior (Abraham & Michie, 2008; Bartholomew, Parcel, Kok, Gottlieb, & Fernández, 2011; Craig et al., 2008). This knowledge is essential for understanding how to design complex interventions to induce change. In addition, there is a growing need for an evidence-based instrument that can be used to evaluate and qualify the potential of existing interventions (e.g., Brug et al., 2010; Craig et al., 2008).

Various meta-analyses have shown that interventions targeted on health behavior change may be effective (Kroese, Werhkan, & Brug, 2006; Shahab & McEwen, 2009). These studies have difficulties in identifying which particular BCTs are responsible for heterogeneity in effectiveness of interventions. An exception is the study by Albarracín et al. (2005), which made an attempt to identify theoretically derived BCTs within HIV prevention programs and showed that some technique types were more likely to effectively change behavior than others, increasing our understanding of why variation in intervention effectiveness exists. In general, however, comparisons between BCTs in intervention effectiveness studies and meta-analyses have been hampered by the lack of a systematic framework for identifying BCTs within interventions.

Recent developments of taxonomies of BCTs provide frameworks that can be used to classify interventions in a systematic way. As such, they provide the possibility to systematically evaluate theory-based BCTs within complex interventions (Abraham & Michie, 2008). A BCT taxonomy clarifies differences and similarities in content of interventions targeting similar behaviors in similar settings. It provides a detailed definition of each BCT, including essential elements. In the taxonomy of Abraham and Michie (2008), for example, Prompt intention formation is defined as “encouraging the person to decide to act or set a general goal” (see Table 1 for an overview of the taxonomy).

Some meta-analyses that have used a BCT taxonomy have succeeded in identifying BCTs that influence the effectiveness of interventions. Webb, Joseph, Yardley, and Michie (2010) examined the effectiveness of Internet-based interventions using a taxonomy adapted from Hardeman, Griffin, Johnston, Kinmonth, and Warham (2000). They found that the two BCTs that were associated with the greatest changes in behavior were Stress management and General communication skills training. Moreover, they found that intervention effectiveness was larger when more BCTs were included. Michie, Abraham, Whittington, McAteer, and Gupta (2009) examined the effectiveness of physical activity (PA) and healthy eating (HE) interventions using the taxonomy of 26 BCTs from Abraham and Michie (2008). They showed that interventions were most likely to be effective when Self-monitoring was used as a technique, or when Self-monitoring plus an additional self-regulation technique were used. Using the same taxonomy, Dombrowski et al. (2012) identified several BCTs (including Self-monitoring) with greater probability of intervention success on weight and kilocalorie consumption. In addition, Dombrowski et al. (2012) identified several BCTs that hampered intervention success, including Provide general information and Provide information on consequences. De Bruin, Vichthauer, Hespers, Schaalma, and Kok (2009) also used the same taxonomy to code standard care in control groups of studies evaluating the effectiveness of interventions for HIV. They showed that the control groups differed greatly in number and type of BCTs. Finally, refined taxonomies have been used to successfully identify BCTs that increase effectiveness for reducing excessive alcohol consumption (e.g., Michie et al., 2012).

When interventions use multiple BCTs, several situations may occur: (a) the effects of the BCTs are additive, (b) the effects of the BCTs cancel out, or (c) the effects of the BCTs amplify. This latter effect is the focus of our study. The amplification of effects implies that a combination of BCTs has a synergistic effect, which is also called an interaction effect. A synergistic effect occurs if the combination of two or more BCTs has a more potent effect than would be expected by their additive effect. For instance, from literature on fear appeals, it has been suggested that Fear arousal as a strategy can only be effective when also Skill information is provided (Rogers, 1995; Ruiters, Abraham, & Kok, 2001). Thus, Fear arousal and Skill information do not enhance success when applied separately, but their combined use can be quite effective (Peters, Ruiters, & Kok, 2013). Similarly, Implementation intentions have been suggested to be effective only when people are sufficiently motivated to engage in specific behavior (Sheeran, Webb, & Gollwitzer, 2005). Thus, Implementation intentions have to be combined with a motivation-enhancing technique to achieve success. Generally, it is expected that BCTs have synergistic effects (Malotte et al., 2000; Michie et al., 2009; Rothman, Baldwin, & Hertel, 2004), and it is considered to be important to gain an understanding on which combinations of BCTs matter (Dixon & Johnston, 2010; Michie et al., 2009). In addition, insight into the combination of techniques is essential with regard to the development of interventions. The synergistic effects of BCTs, however, generally cannot be examined by means of meta-analysis due to the lack of power in meta-regression to identify interaction effects (e.g., Michie et al., 2009). As such, only univariate or, to a lesser extent, additive effects have been examined.

In the present study, the use of classification and regression trees (CART; Breiman, Friedman, Olshen, & Stone, 1984) has been proposed to identify synergistic effects. CART is especially suitable for data with many predictor variables that could interact. CART has been used in the field of health psychology and medical sciences, for example, to examine which combination of factors can predict cancer (van Dijk, Steyerberg, Stenning, & Habbema, 2004), or to stratify patients on disease severity (Trujillano, Badia, Serví, March, & Rodriguez-Pozo, 2009). As far as known, CART has never been used in the field of meta-analysis, except for a small study by Dusseldorp (2001). The aim of the present study was to gain a further understanding of synergistic effects of BCTs by applying CART in a special way to meta-analytic data. This novel approach will be referred to as Meta-CART. The main objective was to examine which combinations of BCTs explain intervention success.

### Method

Data from the 101 studies that were included in the meta-analysis of Michie et al. (2009) were used. The studies were published between 1990 and 2008 in peer-reviewed journals written in English. The study effect size data and the scores on the taxonomy of 26 BCTs were obtained from the authors of the meta-analysis.
Characteristics of Interventions Used in Included Studies

The included studies reported on interventions targeted at adults (18 years and older) to increase their levels of PA or HE. In addition, the studies used an experimental or quasi-experimental design, and applied cognitive or BCTs. Examples of interventions were PA interventions (e.g., Harland et al., 1999), nutrition education (Oenema, Tan, & Brug, 2005), and interactive computer-tailored interventions for PA and HE (Vandelanotte, De Bourdeaudhuij, Sallis, Spittaels, & Brug, 2005). The target population in the studies varied from the general population to patients at spe-
cific risk (e.g., at risk of cardiovascular disease). In order to evaluate the intervention effect, all studies compared an intervention condition with a control or standard-care condition. The 101 studies reported on the effects of a total of 122 interventions, of which 69 were targeted at PA and 53 were targeted at HE. In this study, the interventions have been taken as the analytic level. Following Michie et al. (2009), the PA and HE interventions were considered together, because Michie et al. showed that PA and HE interventions had similar mean effect sizes (i.e., 0.32 and 0.31, respectively).

Outcome Measures

All studies used objective or validated self-reported outcome measures (Michie et al., 2009). In cases in which multiple outcome measures were reported for one evaluation, the following decisions were made: (a) for PA evaluations, the first of the following sequence was selected: exercise level, energy expenditure, percent active, body mass index; and (b) for HE evaluations, the first of the following sequence was selected: diet score, food intake, fat intake, fruit and vegetables/fiber, fruit or vegetables/fiber. Study effect sizes were computed as the standardized mean difference, with a correction for small sample size, Hedges’s g (Hedges, 1981). For the purpose of this study, the distribution of the 122 study effect sizes was dichotomized using a median split (also see the Meta-CART section). The median was at 0.31, which can be regarded as a small to moderate effect size in the area of behavioral sciences (Cohen, 1988). In addition, the overall pooled effect size found by Michie et al. (2009) was 0.31, which can be regarded as the average effect that can be achieved by HE and PA interventions. Therefore, we used this value as a criterion for success for this type of interventions. Interventions with an effect size above 0.31 were classified as more successful (coded with a 1), and those with effect sizes below or equal to 0.31 were classified as less successful (coded with a 0). As a result, these two categories contained 61 interventions each.

BCTs

Each intervention was coded by Michie et al. for inclusion or not of each of the 26 BCTs from the taxonomy of Abraham and Michie (2008). In line with the approach of Dixon and Johnston (2010), the BCTs were grouped into categories referring to their motivational and self-regulatory functions derived from several theories (Austin & Vancouver, 1996; Gollwitzer, 1996; Heckhausen, Wrosch, & Schulz, 2010; Rothman et al., 2004). The following three categories of BCTs were distinguished: (a) motivation enhancing, (b) planning and preparation, and (c) goal striving and persistence.

BCTs were considered motivation enhancing when aimed at the motivational aspect: working toward an intention by influencing its determinants (such as risk perception and outcome expectancies), for example, Provide information on consequences. The planning and preparation category included techniques that assume action preparation, such as Provide instruction. The goal striving and persistence category included techniques that are aimed at continuation and evaluation of behavior and prevention of relapse, for example, Provide feedback on performance. Although a specific technique may contribute to various self-regulatory processes, the technique was grouped within the category for which it was expected to be most important. The grouping of the BCTs was performed by authors LvG and PvE separately. Differences were resolved through discussion. Table 1 gives an overview of the grouping and, for each BCT, the number of interventions that included the technique. Four techniques present in the taxonomy were excluded in this study because fewer than five interventions used these techniques (see techniques mentioned in note of Table 1). In this case, subgroup analysis is not feasible. The techniques most frequently used were Prompt intention formation, Provide instruction, and Provide information on consequences (see Table 1). Among the remaining 22 techniques, the interventions included an average of six techniques, ranging from none (Hivert, Langlois, Berard, Currier, & Carpentier, 2007) to 13 (Burke, Giangiulio, Gillam, Beilin, & Houghton, 2003).

CART

CART is a machine learning technique that builds classification and regression trees. Classification trees are used to model categorical outcome variables, whereas regression trees model continuous outcome variables. The CART algorithm partitions subjects (or interventions in our case) into more homogeneous subsets, resulting in a binary tree in which the end nodes contain the most detailed groups. Figures 1, 2, 3, and 4 give examples of classification trees that distinguish more successful interventions from less successful ones, using the dichotomized effect sizes as the outcome variable. CART starts by placing all interventions into one group (i.e., the root node). Then it creates two groups (i.e., child nodes) by tentatively splitting the observations below and above a chosen cut point on a chosen predictor variable (e.g., a particular BCT), and records how well this grouping (also called partitioning) predicts the dichotomized outcome variable (e.g., more or less successful, coded with $Y = 1$ or $Y = 0$). This “how well” is defined in terms of a partitioning criterion. For classification trees, the Gini index is most frequently used as partitioning criterion (Breiman et al., 1984). In the case of an outcome with two categories, the Gini index in a node $t$ equals 2 times the product of the probabilities of each of the two categories: $2p(Y = 1)p(Y = 0)$. It can be seen as a measure of heterogeneity of the outcome in a node. Its maximum value is 0.50 and implies, in our application, that half of the studies in a node are more successful (i.e., $Y = 1$), and half of the studies are less successful (i.e., $Y = 0$). A node is more homogeneous if its Gini index is lower. The splitting process is repeated for all possible cut points and all predictors, and the best split in terms of the partitioning criterion (e.g., the lowest Gini index) is used to grow the tree. The process repeats itself on the child nodes. For example, in the tree in Figure 2, Technique 4 (Prompt intention formation) is selected for the first split. If this technique is not used, interventions belong to the left child node. If the technique is used, interventions belong to the right child node. Each of the child nodes is then a candidate node for the next split. In our example tree (see Figure 2), the right child node is split into two child nodes using Technique 1 (Provide information about behavior–health link) as splitting variable. The end nodes, also called leaves (i.e., squares in Figure 2), are the final subgroups. In the tree of Figure 2, there are no perfectly homogeneous subgroups (e.g., a node that contains only successful interventions). Going from the upper left to the upper right end
node, the percentage of interventions within a node that are successful (i.e., above the threshold of 0.31) is 36%, 41%, and 77%, respectively.

**Meta-CART**

During initial analyses, we encountered the following disadvantages of applying CART directly to the study effect sizes: (a) the solution of CART, a regression tree, was very unstable (shown by cross-validation results); (b) the sample sizes of studies were not taken into account (CART has no possibility to weight the outcome data); and (c) CART has no possibility to distinguish between random and fixed effects. To overcome these disadvantages, we used a two-step strategy. In a first step, CART was fitted using the dichotomized effect sizes as outcome. The resulting classification tree appeared to be more stable than the regression tree, suggesting that the Gini index was a more stable partitioning criterion (also see Hastie, Tibshirani, & Friedman, 2001). In a second step, a standard subgroup meta-analysis was performed using all original effect size measures (gs) as outcome variable. We use the term Meta-CART to refer to this approach and will explain the steps in more detail.

The CART analysis of the first step of Meta-CART was performed in a standard way, which consisted of two parts. First, a large tree was grown, using a minimum of five interventions in an end node and a minimal decrease in heterogeneity (impurity) of 0.001 as stopping rules (Breiman et al., 1984). Second, tenfold cross-validation was performed to avoid overfitting of the data. The best tree size (i.e., the number of end nodes) was selected as the smallest one that satisfied the one-standard-error rule (Breiman et al., 1984). This means that its cross-validated error was smaller than the minimum cross-validated error plus one standard error. To increase the stability of the results, the cross-validation procedure was repeated 1,000 times. This resulted in 1,000 estimates of the best tree size, from which the mode or median was chosen as final estimate of best tree size. The cross-validation procedure ensures that the data are not overfitted (Hastie et al., 2001); in other words, it ensures that the final tree (i.e., the synergistic effect) can be generalized to future observations. The final tree represents a synergistic effect between the BCTs that are used as splitting variables. The end nodes of the tree form the subgroups. From a tree, a new variable was created, with its categories referring to the end nodes of the tree. For example, for Figure 2, the interventions in the nodes from left to right, receive a value of 1 to 3 on this new grouping variable, respectively.

![Figure 1. Classification tree across all categories for those studies that used at least one technique (n = 120). Plots in the end nodes display the percentage of interventions that were more successful (i.e., an effect size higher than 0.31).](image-url)
The subgroup meta-analysis of the second step of Meta-CART was performed using a mixed effects model to investigate whether the new grouping variable resulting from the first step accounted for the heterogeneity in the study effect sizes. The mixed effects model consisted of a random effects model within subgroups and a fixed effect model across subgroups, which is an approach recommended by Borenstein, Hedges, Higgins, and Rothstein (2009). The $p$ value of the between-groups $Q$ statistic indicated whether the grouping effect was significant (Borenstein et al., 2009). A two-sided significance level of .05 was used. An advantage of the subgroup analysis was that a mean effect size (weighted by sample size) was obtained for each subgroup (i.e., end node of the tree). The mixed effects analysis was performed in Comprehensive Meta Analysis, version 2.2 (Borenstein, Hedges, Higgins, & Rothstein, 2005).

**Analysis Strategy**

Prior to the performance of the analyses, the study effect sizes (gs) were inspected for outliers. Analyses were performed with and without these outlier(s). In total, four Meta-CART analyses were performed, varying in the BCTs that were included as predictor variables in the analyses. In the first analysis, all BCTs were included as predictor variables (except the four BCTs that were used by too few interventions; see Table 1). The latter three analyses were performed to investigate whether synergistic effects within a behavior change category were present. Only those studies that used at least one technique from the set of selected BCTs were included in an analysis.

**Results**

The study effect sizes (gs) ranged from −.17 to 1.90, with an overall effect size (weighted for sample size) of 0.31 (95% CI [0.26, 0.36]; also see Michie et al., 2009). The oldest study (Insull et al., 1990) was an outlier in terms of effect size ($g = 1.90$). Without this study, the maximum observed effect size ($g$) was 1.28. The results presented are from the analyses without the outlier. Details about differences in results with and without the outlier will be given in the Discussion and Conclusions.

The Meta-CART analysis including all BCTs as predictor variables (22 in total) resulted in a classification tree with eight end nodes (see Figure 1). The subgroup analysis showed that the difference between the eight groups in mean effect sizes was significant ($p = .02$; Table 2). A synergistic effect was found between Prompt review of behavioral goals, Provide information on consequences, Provide information about behavior–health link, and Use follow-up prompts (left side of Figure 1). Less effective were those interventions that did not use Prompt review of behavioral goals, but used Provide information on consequences without Provide information about behavior–health link (mean effect size, $g = 0.23$; Table 2, Group 1). Most effective were those interventions that did not use Prompt review of behavioral goals, but used Provide information on consequences in combination with Provide information about behavior–health link and Use follow-up prompts ($g = 0.44$; Table 2, Group 3). Another synergistic effect
was found between Prompt review of behavioral goals, Provide information on consequences, Provide feedback on performance, Provide instruction, and/or Provide contingent rewards (right side of Figure 1). Least effective were those interventions without Prompt review of behavioral goals, without Provide information on consequence, with Provide feedback on performance, but without Provide instruction ($\bar{g} = 0.05$; Table 2, Group 4). More effective than average were those interventions that used Prompt review of behavioral goals ($\bar{g} = 0.40$; Group 8).

The Meta-CART analysis including the motivation-enhancing BCTs as predictor variables (four in total; Table 1) resulted in a classification tree with three end nodes (see Figure 4). Subgroup analysis showed that the difference between these groups in mean effect sizes was not significant ($p = .08$; Table 2).

Discussion and Conclusions

This reanalysis of data from a systematic review of PA and HE interventions (Michie et al., 2009) aimed to identify synergistic effects of BCTs. Meta-CART was applied as a novel analysis strategy for subgroup analysis within meta-analysis. The BCTs were grouped into three categories of behavior change: motivation enhancing, planning and preparation, and goal striving and persistence. Four Meta-CART analyses were performed: One analysis included all BCTs together (22 in total) as predictor variables, and the other three included the BCTs from one category. Meta-CART identified several combinations of techniques that were more likely to effectively change behavior and combinations of techniques that were less likely to successfully change behavior. The results from the analysis with all BCTs revealed that the combination of Provide information on the consequences, Provide information about behavior–health link, and Use follow-up prompts was most successful in achieving behavior changes. In addition, the use of Provide feedback on performance, without the use of Prompt review of behavioral goals, Provide information on the consequences, and Provide instruction was least successful in achieving behavior changes. From the motivation-enhancing category, the combination of Prompt intention formation and Provide information about behavior–health link was found to be most successful in achieving behavior changes. No significant synergistic effects were found for the BCTs from the planning and preparation category or the goal striving and persistence category separately.

Our findings showed that the strongest synergistic effect was found with motivation-enhancing BCTs. Of particular interest was the fact that those interventions that included Prompt intention formation, but did not use Provide information about behavior–health link, showed the lowest mean effect size ($\bar{g} = 0.24$) in this category. This finding seems to suggest that those interventions that aim to motivate change, without addressing the perceived need for changing (e.g., personal susceptibility), are actually worse off than the average intervention effect (i.e., $\bar{g} = 0.31$). This result seems to be in accordance with theories—such as the precaution adoption process model (Weinstein, Rothman, & Sutton, 1998)—that highlight the fact that perceived vulnerability is a prerequisite for actually thinking about change. Our findings also show that interventions that use the combination of Prompt intention formation and Provide information about behavior–health link showed the highest mean effect size ($\bar{g} = 0.46$). This finding puts the result of Dombrowski et al. (2012)—who found that interventions including Provide information about behavior–health link (T1) were less successful—in another perspective. Apparently, it is important to combine the two BCTs instead of using them separately. This result could not be found by the often applied univariate approach to subgroup meta-analysis, which was used by Dombrowski et al.

Across the three behavior change categories, three results were striking. First, Prompt review of behavioral goals appeared to be an important predictor of intervention success. The 19 interven-

Figure 4. Classification tree for the goal striving and persistence category for those studies that used at least one technique from this phase ($n = 108$). Plots in the end nodes display the percentage of interventions that were more effective (i.e., an effect size higher than 0.50). The subgroup analysis showed that the difference between these groups in mean effect sizes was not significant ($p = .08$; Table 2).
also used several other techniques. For example, 12 of the 19 interventions successful. The 19 interventions from this subgroup also used concluding that interventions using only this technique will be calculating such self-regulatory strategies indeed seem beneficial for choices (Heckhausen et al., 2010).

Prompt review of behavioral goals. The combination is a useful alternative for the successful technique persistence category (Gollwitzer & Sheeran, 2006). Third, those interventions that used Provide feedback on performance as a technique, without using Provide instruction, Provide information on consequences, and Prompt review of behavioral goals, were least effective ($\bar{g} = 0.05$). Of these latter three, the lack of Provide instruction seemed important, because those interventions that used the combination of Provide feedback on performance and Provide instruction, without the use of the other two BCTs, showed an effect size similar to the average ($\bar{g} = 0.31$). These results suggest that providing feedback on performance may have a counterproductive effect when not providing clear instruction of the behavior. This effect may occur when people lack goal commitment, feel incapable of making a change, or simply are unaware of the opportunities to perform (e.g., Ajzen, 1991; Gollwitzer & Sheeran, 2006; Latham & Locke, 1991; Oettingen, Pak, & Schnetter, 2001).

Table 2

<table>
<thead>
<tr>
<th>Group</th>
<th># interv.</th>
<th>$\bar{g}$</th>
<th>95% CI</th>
<th>Q (df)</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>32</td>
<td>0.23</td>
<td>0.15, 0.31</td>
<td>16.8 (7)</td>
<td>.02</td>
</tr>
<tr>
<td>Group 2</td>
<td>17</td>
<td>0.30</td>
<td>0.20, 0.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 3</td>
<td>6</td>
<td>0.44</td>
<td>0.27, 0.61</td>
<td>25.2 (2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Group 4</td>
<td>7</td>
<td>0.05</td>
<td>-0.11, 0.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 5</td>
<td>12</td>
<td>0.31</td>
<td>0.20, 0.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 6</td>
<td>8</td>
<td>0.26</td>
<td>0.06, 0.45</td>
<td>6.1 (2)</td>
<td>.11</td>
</tr>
<tr>
<td>Group 7</td>
<td>19</td>
<td>0.32</td>
<td>0.24, 0.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 8</td>
<td>19</td>
<td>0.40</td>
<td>0.26, 0.55</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: The end nodes of the trees (Figures 1 to 4) were the categories of the grouping variable, and the study effect sizes were used as outcome variable. The first column (Group) represents the numbering of the end nodes of each tree from left to right. Results are shown excluding the outlying effect size from Insull et al. (1990).

$\#$ interv. = number of interventions included in the subgroup; $\bar{g}$ = average effect size, weighted for sample size; CI = confidence interval; $Q = \text{Between groups } Q$ statistic; df = degrees of freedom.
plained the greatest amount of intervention success. In contrast, our findings showed that Reviewing behavioral goals was a more influential technique. This difference can be explained by the treatment of the Insull et al. (1990) intervention, which, in our study, was considered to be an outlier. Analyses including this intervention did not greatly alter the CART findings, given that these analyses did not depend on the size of the effect. Inclusion of the Insull et al. study, however, had a large effect on the meta-analytic results (i.e., the subgroup analyses). The weighted average effect size of the subgroup with the intervention of Insull et al. (that used Self-monitoring as a behavior change technique) was much higher than without this intervention (i.e., 0.42 vs. 0.34).

Important to note is the fact that the univariate subgroup analysis of Michie et al. could not identify any significant behavior change technique, whereas our subgroup analyses using combinations of techniques could. This latter finding confirms the conjecture made by Michie et al. (2009), who stated that it is likely that combinations of BCTs may interact to account for effect size heterogeneity (p. 698).

Some limitations of the present study need to be acknowledged. First, our analyses were based on the taxonomy of Abraham and Michie (2008); four BCTs of this taxonomy were not included in the CART analyses, given the limited number of studies that applied these particular techniques. In addition, other techniques that are not part of the taxonomy could be important. Moreover, it is plausible that other study features may be responsible for heterogeneity of the effect sizes. Michie et al. (2009) showed, however, that differences in the effectiveness of the interventions could not be explained by, among others, duration of the intervention, format of delivery (e.g., group or individual), setting, and target population. Furthermore, we linked the BCTs to three categories of self-regulatory processes (e.g., Gollwitzer & Sheeran, 2006; Rothman et al., 2004). Of note, other frameworks of behavior change exist, for example, the COM-B system, which emphasizes the following conditions for behavior: capability, opportunity, and motivation (Michie, Van Stralen, & West, 2011). This framework would consider another grouping of the BCTs (e.g., education, persuasion, incentivization, training, and enablement; see Michie et al., 2011). Hence, an open question remains as to whether a different grouping of the techniques would have resulted in other synergistic effects.

Finally, two other important caveats of the present study should be addressed when considering the development and evaluation of complex interventions. First, the study provides information on which combination of techniques could enhance the effectiveness of PA and HE interventions. It does not, however, explain how the techniques should be used in practice. As pointed out by several studies (Bartholomew et al., 2011; Kok, Schaalm, Ruiter, Van Empelen, & Brug, 2004; Rothman et al., 2004), it is important to understand the context in which a behavior change technique operates. Second, our study shows the effects on the actual behavior outcomes, but not on the potential mediators (e.g., factors such as attitudes, self-efficacy and skills) that might explain behavior change (Albarracin et al., 2005; Webb & Sheeran, 2006). Given that behavior change is assumed to be the result of changes via behavioral or environmental determinants (Kremers et al., 2006; Michie et al., 2011), and BCTs are expected to have an impact on such determinants, it is important to acquire more insight into the change mechanisms (e.g., Olander et al., 2013).

Our study illustrates that Meta-CART is useful in identifying synergistic effects that cannot be found by the more widely used approach of meta-regression and subgroup meta-analysis. Meta-CART combines CART and meta-analysis and has some particular advantages: (a) it has been shown to be more powerful in identifying combined effects of BCTs, (b) it is less sensitive to potential outliers, (c) it enables the weighting of studies’ effects, and (d) it provides meaning (in terms of average effect sizes) to differences in effect. In future meta-analyses that investigate synergistic effects, the use of Meta-CART is recommended.

In summary, this study provides evidence for effective combinations of BCTs. Particular combinations that were more successful than average (e.g., Provide information about behavior—health link with Prompt intention formation) were identified, as were particular combinations that hampered success (e.g., Provide feedback on performance without Provide instruction). The understanding and empirical grounding of combinations of effective techniques is likely to contribute to the improved design and evaluation of interventions.

References


Received December 11, 2012
Revision received June 10, 2013
Accepted July 25, 2013

Members of Underrepresented Groups: Reviewers for Journal Manuscripts Wanted

If you are interested in reviewing manuscripts for APA journals, the APA Publications and Communications Board would like to invite your participation. Manuscript reviewers are vital to the publications process. As a reviewer, you would gain valuable experience in publishing. The P&C Board is particularly interested in encouraging members of underrepresented groups to participate more in this process.

If you are interested in reviewing manuscripts, please write APA Journals at Reviewers@apa.org. Please note the following important points:

- To be selected as a reviewer, you must have published articles in peer-reviewed journals. The experience of publishing provides a reviewer with the basis for preparing a thorough, objective review.
- To be selected, it is critical to be a regular reader of the five to six empirical journals that are most central to the area or journal for which you would like to review. Current knowledge of recently published research provides a reviewer with the knowledge base to evaluate a new submission within the context of existing research.
- To select the appropriate reviewers for each manuscript, the editor needs detailed information. Please include with your letter your vita. In the letter, please identify which APA journal(s) you are interested in, and describe your area of expertise. Be as specific as possible. For example, “social psychology” is not sufficient—you would need to specify “social cognition” or “attitude change” as well.
- Reviewing a manuscript takes time (1–4 hours per manuscript reviewed). If you are selected to review a manuscript, be prepared to invest the necessary time to evaluate the manuscript thoroughly.

APA now has an online video course that provides guidance in reviewing manuscripts. To learn more about the course and to access the video, visit http://www.apa.org/pubs/authors/review-manuscript-ce-video.aspx.