Health-Promotion Interventions Targeting Multiple Behaviors:  
A Meta-Analytic Review of General and Behavior-Specific Processes of Change

Angela L. Zhang & Sicong Liu
University of Pennsylvania

Benjamin X. White
University of Illinois

Xi C. Liu, Marta Durantini, Sally M-p Chan, Wenhao Dai, Yubo Zhou, Melody Leung, Qijia Ye, Devlin O’Keefe, Lidia Palmese, & Dolores Albarracín
University of Pennsylvania

Author Note

Angela L. Zhang and Sicong Liu contributed equally and share the 1st authorship. An original version of this work was part of Benjamin X. White’s dissertation. The data supporting the findings reported in this study are openly available in OSF at https://osf.io/qsx57/?view_only=6fffd7080d844e98cbd2d0834ed61e9.

Corresponding Author: Dolores Albarracín, Department of Psychology; University of Pennsylvania (dalba@upenn.edu)

In Press, Psychological Bulletin
Abstract
Although health-promotion interventions that recommend changes across multiple behavioral domains are a newer alternative to single-behavior interventions, their general efficacy and their mechanisms of change have not been fully ascertained. This comprehensive meta-analysis (6,878 effect sizes from 803 independent samples from 364 research reports, $N = 186,729$ participants) examined the association between the number of behavioral recommendations in multiple-behavior interventions and behavioral and clinical change across eight domains (i.e., diet, smoking, exercise, HIV [Human Immunodeficiency Virus] prevention, HIV testing, HIV treatment, alcohol use, and substance use). Results showed a positive, linear effect of the number of behavioral recommendations associated with behavioral and clinical change across all domains, although approximately 87% of the samples included between 0 and 4 behavioral recommendations. This linear relation was mediated by improvements in the psychological well-being of intervention recipients and, in several domains (i.e., HIV, alcohol use, and drug use), suggested behavioral cuing. However, changes in information, motivation, and behavioral skills did not mediate the impact of the number of recommendations on behavioral and clinical change. The implications of these findings for theory and future intervention design are discussed.

*Keywords*: multiple-behavior intervention, efficacy, meta-analysis, mechanism of change, public health
Public Significance Statement:

The present meta-analysis examined the effects of multiple-behavior interventions and processes of change. A higher number of recommendations is associated with more positive behavioral and clinical outcomes. Analyses of potential mediators suggest two key mechanisms of change underlying this relation: (a) cuing of one behavior by another and (b) improvements in psychological well-being. This work advances our knowledge of the optimal number of recommendations to include in health-promotion interventions as well as the mechanisms through which multiple recommendations operate. Thus, it can inform the design of future interventions to increase their efficacy in facilitating positive health outcomes.

OSF Pre-registration: https://osf.io/mq5kr/?view_only=54d04d57b5ef45aaa480a47259cb91cd

OSF Open Data: https://osf.io/qsx57/?view_only=6fffd7080d844e98cbd2d0834ed61e9

Competing interests: The authors declare no competing interests.
Health-Promotion Interventions Targeting Multiple Behaviors:
A Meta-Analytic Review of General and Behavior-Specific Processes of Change

Clusters of unhealthy behaviors contribute to the most prevalent diseases in the United States and most areas of the world (Farnham et al., 2010; Heron, 2019; Prevention, 2018). As a result, the most promising behavioral-health programs have moved from attempting to change a single behavior to attempting to change behavioral clusters such as improving lifestyle (King et al., 2015). For example, interventions targeting both diet and exercise have decreased cardiovascular risk (Salas-Salvadó et al., 2019), delayed onset of type 2 diabetes (Gong et al., 2019), and improved cognitive functioning (Ngandu et al., 2015). Other interventions have been successful in simultaneously decreasing both substance use and risky sexual behaviors (Mouttapa et al., 2009; Tucker et al., 2017). Prior syntheses of the efficacy of these programs (Dai et al., 2020; Sunderrajan et al., 2021; Wilson et al., 2015), however, have looked at a small number of domains (i.e., one to three) and left critical questions unanswered. One of these questions is whether higher numbers of recommendations (i.e., exhortations to perform specific health behaviors) increase behavioral and clinical change across diverse behavioral domains. Whereas the number of recommendations in lifestyle programs has shown to facilitate change only up to a point, the number of recommendations in drug and alcohol programs as well as HIV (Human Immunodeficiency Virus) programs appears to increase clinical and behavioral change in a linear fashion. Therefore, the optimal number of behavioral recommendations across the board and the processes underlying the efficacy of multiple-behavior interventions remain unclear.

Another limitation of the literature is the absence of a process framework about multiple-behavior change mechanisms and evidence testing those mechanisms. In this article, we
introduce a conceptualization in which multiple behavioral recommendations can trigger either general or behavior-specific processes of behavioral change. General processes of behavioral change tend to occur spontaneously and apply to different behavioral domains, thus being relatively efficient. These processes include cuing, a mechanism in which changes in one behavior cue changes in other behaviors (for related processes, see Bargh et al., 1992; Weingarten et al., 2016; Wood, 2017) such as alcohol drinking cuing consumption of calorie-dense snacks (Carels et al., 2014). General processes of change also comprise changes in psychological well-being, including reductions in depression and anxiety symptoms (Worley et al., 2012) and increases in self-control (Hofmann et al., 2014), all of which cut across behaviors. These general processes are likely to occur by simply facilitating multiple-behavior change efficiently via the introduction of new cues to positive behaviors (Albarracín, 2021; Bargh, 1994) or via a reduction in stress and consequent engagement in unhealthy behaviors to reduce stress. Consequently, these processes may lead to interventions with higher numbers of behavioral recommendations exerting a linear impact on change in behavioral and clinical outcomes. Prior work examining the benefits of one behavior for other behaviors has referred to these changes as “spill-over” (Mata et al., 2009) or “transfer” (Lippke et al., 2012) effects.

In contrast, behavior-specific processes of change require thinking about the specific behaviors recommended in an intervention. For example, interventions that make more recommendations may increase information and behavioral skills acquisition, such as knowledge about cardiovascular health and grocery shopping skills, which can in turn facilitate regular exercise and a healthy diet (Kiernan et al., 2013). Moreover, multiple-behavior interventions may improve motivation by strengthening attitudes, norms, and intentions to perform each recommended behavior (Albarracin et al., 2001). Nevertheless, processing information about
each behavior may require cognitive capacity and motivation to think about the behaviors (Albarracín, 2002), thus limiting how many recommendations can be introduced (Wilson et al., 2015). Consequently, if these processes dominate, they may result in a curvilinear effect of the number of behavioral recommendations on behavioral and clinical change due to each recommendation incurring costs that diminish benefits as the number of recommendations increases. Compensatory effects (e.g., indulging in unhealthy snacks now with the intention of exercising more later; Knäuper et al., 2004) have after all been identified in several domains. For example, physically active people consume more alcohol, people who frequently consume alcohol engage in more physical activity (Nigg et al., 2009), and participating in an exercise intervention can lead to increasing energy intake (Martin et al., 2019). Similarly, workers who switch from a sedentary occupation to a more physically active one compensate by exercising less in their leisure time, whereas the opposite pattern is found for workers who switch from a physically active occupation to a sedentary one (Nooijen et al., 2018).

In this study, we conducted a meta-analysis of multiple-behavior interventions to assess the extent to which cuing and increases in psychological well-being, information, motivation, and behavioral skills help to explain the effects of multiple behavioral recommendations on behavioral and clinical change. We first evaluated the shape of the relation between the number of recommendations and intervention efficacy, considering both behavioral outcomes (self-report measures of behavior; e.g., self-report scales) and clinical outcomes (e.g., blood pressure, cholesterol, STI test results). This synthesis included a comprehensive dataset spanning eight prominent behavioral intervention domains: (a) exercise, (b) diet, (c) smoking, (d) HIV prevention, (e) HIV testing, (f) HIV treatment, (g) alcohol use, and (h) substance use. These domains are of particular interest to behavioral change researchers because they represent some
of the leading modifiable risk factors associated with the highest burdens of disease, death, and economic costs both in the United States and globally (Bolnick et al., 2020; Mokdad et al., 2018; Murray et al., 2020; Vos et al., 2020), and the literatures that cover them have been used to test prior models of behavioral change (for examples, see Carpenter, 2010; McEachan et al., 2011). We thus examined behavioral cuing and changes in psychological well-being, information, motivation, and behavioral skills in these diverse and important areas.

**Single-Behavior Interventions and Their Shortcomings**

The first generation of interventions developed to change behavior focused on changing one behavior deemed important for a population. Early HIV prevention interventions in the 1980s, when the disease was first identified, focused on increasing condom use (Wolitski, Janssen, et al., 2005), and many alcohol and substance use treatment programs have focused on only one problem behavior (Murthy, 2016). Weight loss interventions have generally been more mixed in that some recommend both exercise and diet (e.g., Parsons et al., 2005), but many still recommend solely changes in either diet or physical activity, with early meta-analyses noting that about 70 percent of studies recommended a single behavior as the target for change (Conn et al., 2002).

One limitation of single-behavior interventions has been that they routinely produce only small to medium effect sizes in behavioral or clinical change ($d = 0.3$, interpreted as a change of 0.3 standard deviations). For example, a meta-analysis of interventions to increase physical activity found that the average intervention changed physical activity by about a quarter of a standard deviation ($d = 0.27$; R. E. Rhodes et al., 2017). Interventions to curb alcohol ($d = 0.28$; Dotson et al., 2015); quit smoking ($d = 0.34$; Wu et al., 2006); and prevent HIV ($d = 0.40$; Albarracin et al., 2005) have all shown medium effect sizes as well.
The limitations of single-behavior interventions are perhaps not surprising given the reality of the challenges people experience with respect to changing their health behaviors. Several observational studies noted that, rather than people performing a single unhealthy behavior, they often have clusters of unhealthy practices (Keller et al., 2011; Lippke et al., 2012; Ritchwood et al., 2015). For lifestyle behaviors, there are positive correlations between exercise and diet: \( r = .16 \) to \( .26 \); between exercise and smoking: \( r = .21 \); and between diet and smoking: \( r = .11 \) (Keller et al., 2011; Lippke et al., 2012). A meta-analysis of observational studies of adolescent sexual health behaviors also found moderate associations between use of alcohol and/or substances and risky sexual behaviors: \( r = .22 \); between alcohol use and unprotected sex: \( r = .11 \); between substance use and unprotected sex: \( r = .18 \); between alcohol use and multiple partners: \( r = .25 \); and between substance use and multiple partners: \( r = .25 \) (Ritchwood et al., 2015). What these observational studies suggest is that single-behavior interventions may be ill equipped to address these patterns of correlated behaviors.

Multiple-Behavior Interventions

Noting the limitations of single-behavior interventions, behavioral change researchers began developing and testing multiple-behavior interventions (Noar et al., 2008; Prochaska et al., 2008; Smedley & Syme, 2001). A multiple-behavior intervention is characterized by recommending change in at least two health-behavior domains either simultaneously or sequentially. Where a single-behavior intervention may recommend that clients increase their physical activity (one behavioral domain), a multiple-behavior intervention may recommend changes in lifestyle, such as increasing physical activity, reducing fat intake, and quitting smoking (three behavioral domains). Initially, interventions targeted multiple behaviors because populations had multiple risk factors, expecting a similar amount of change in each behavior and
expecting additive effects on clinical outcomes such as improvements in cardiovascular health (Hjermann et al., 1981; Kjelsberg, 1982; Wilhelmsen et al., 1986). Yet, other studies, such as Spring et al. (2009), have shown that combining smoking and weight control advice not only improves cardiovascular health but also leads to greater smoking abstinence and less weight gain compared to recommending smoking cessation alone. Likewise, interventions that combine drug and alcohol risk reduction with sexual risk reduction lead to larger decreases in sexually transmitted infections than do interventions targeting only sexual risk reduction (Bryan et al., 2018).

Initial tests of whether recommending change in multiple behaviors leads to greater behavioral change for each recommended behavior have been promising. Across three meta-analyses, multiple-behavior interventions outperformed single-behavior ones in how much behavioral change they produced on average (Dai et al., 2020; Sunderrajan et al., 2021; Wilson et al., 2015). Still, the shape of the impact of the number of behavioral recommendations on average clinical and behavioral change remains a question. Both Dai et al. (2020) and Sunderrajan et al. (2021) noted a positive, linear effect of the number of recommendations, such that each recommendation added about $d = 0.07$ of improvement in behavioral and clinical outcomes on average. In contrast, Wilson et al. (2015) noted that studies with two to three recommendations outperformed those with only one ($d = 0.33$ vs. $d = 0.17$), but that there was no difference between single- and multiple-behavior programs once there were four or more recommendations ($d = 0.33$ vs. $d = 0.19$). Yet, these prior syntheses have concerned relatively circumscribed domains, whereas our meta-analysis tried to clarify whether the number of recommendations exerts linear or curvilinear effects on behavioral and clinical change across eight domains of public health significance.
Theoretical Framework: General and Behavior-Specific Processes of Behavioral Change

Despite this important body of past literature, prior meta-analytic work has predominantly focused on demonstrating the efficacy of multiple-behavior interventions in specific health areas rather than examining the likely mechanisms of behavioral change across health behaviors. To fill this gap, we propose that multiple recommendations can operate through either general or behavior-specific processes of behavioral change by which changes in one behavior potentiate changes in other behaviors.

General Processes

General processes of behavioral change are those that generalize across behavioral domains because they do not involve thinking about the specific content of a behavioral recommendation. As a result, these processes require minimal cognitive capacity and motivation to implement the recommended changes and can thus be executed efficiently. One example of a general process is behavioral cuing. Prior research on classical conditioning (Pavlov, 1927) and priming (Bargh et al., 1992; Dijksterhuis et al., 2000) has demonstrated that certain stimuli can provide automatic cues for behavior. More recently, research has shown that habitual behaviors (i.e., habits), defined as behaviors that occur with a high level of automaticity (i.e., low or no cognitive effort), are tied to consistent environmental cues (Verplanken & Orbell, 2003; Wood & Neal, 2007). For example, people reflecting on their behavior report going to the bathroom to brush their teeth with no thought once they start their bedtime routine (McCloskey & Johnson, 2019). Behavioral cuing may occur as a result of habit strength, as habituation of one behavior (e.g., exercise) may free up cognitive resources needed to self-regulate another behavior (e.g., eating) (Fleig et al., 2011). In line with this hypothesis, exercise has been found to be associated with healthy eating habits via increases in exercise habit strength (Fleig et al., 2014; Fleig et al.,
All of these studies thus hint at the possibility that changes in one behavior, such as exercise, may seamlessly produce changes in other behaviors, such as alcohol use.

Changes in psychological well-being (e.g., decreases in stress, anxiety, and depression) represent another general process of behavioral change. Prior research (e.g., Appelhans et al., 2012; Thoits, 2010) has shown that experiencing low psychological well-being (e.g., anxiety, stress, or depression) increases risky behaviors such as binge eating, alcohol use, substance use, and risky sex. Conversely, engagement in unhealthy behaviors, such as smoking, has been implicated as a risk factor for psychological disorders (Firth et al., 2020). Accordingly, a higher number of behavioral recommendations may produce behavioral change through increases in psychological well-being. For example, increases in physical activity have been found to decrease anxiety and depression (Ashdown-Franks et al., 2020; Marcolina, 2007), and antidepressant treatment has been shown to improve exercise performance (Özerbil et al., 2006). More generally, behavioral activation has been shown to lift depression, possibly through dopaminergic mechanisms (Salamone et al., 2016) that may come into play when multiple behaviors are recommended.

In the case of multiple-behavior interventions, general processes would suggest a linear relation between the number of recommendations in an intervention and resulting behavioral or clinical change. For the proposed mechanism of behavioral cuing, this premise is supported by multitasking research findings that people swapping from one task to another (a scenario similar to doing one thing and being cued to do another) perform at the same level when the task swaps automatically and when they are told to prepare to swap tasks (Rogers & Monsell, 1995). Furthermore, the cost of swapping remains stable regardless of how many different behaviors are being swapped (Cho et al., 2015; Kiesel et al., 2010), again implying automaticity in performing
multiple tasks. As a result, executing multiple behaviors may occur without posing significant attentional demands and lead to linear effects of the number of recommendations on behavioral change. The same might be the case for changes in emotional well-being, as promoting well-being typically frees cognitive resources and should influence other behaviors in an effortless fashion. Nevertheless, changes in well-being should be accompanied by changes in stress and/or depression, whereas cuing implies that changes in one behavior mediate changes in another behavior.

**Behavior-Specific Processes**

Behavior-specific processes are likely to involve changes in information, motivation, and behavioral skills pertinent to each behavior (Fisher et al., 2006; Fisher & Fisher, 1992). Thus, these specific processes are likely to require cognitive capacity, motivation, and time because they necessitate thinking about the specific content of a behavioral recommendation. Conscious and effortful processes have been a consistent predictor of behavioral change in multiple areas. Studies of student success, for example, have shown that students who are more engaged with lecture material perform better in both remembering material and applying it properly (Bakker et al., 2015; Robinson, 2013). Furthermore, models of goal pursuit have suggested that tasks that require effort may enhance pursuit and resulting change, showing benefits to behavior-specific processes of change (Bandura, 1977; Lee et al., 2015).

One limitation of behavior-specific processes of behavioral change is that effortful processes can lead to underperformance when resources are overtaxed. For example, when learning new information, confusion and conflict stemming from receiving too much information can reduce responsiveness (Harkins & Petty, 1981) and, of course, recall (Hall et al., 2015; McDonald et al., 2017). People may also devalue hard-to-implement behaviors (Lee et al., 2015)
or burn out when their resources are challenged (Angrave & Charlwood, 2015; Irie et al., 2001). The primary implication of these findings is that recommending a higher number of behaviors may decrease change for each behavior, resulting in a curvilinear relation in which change flattens after a certain number of recommendations is made.

The impact of behavior-specific processes of change is likely to be apparent in changes in information (Fisher et al., 2006; Fisher & Fisher, 1992). Models of health behavior have consistently argued that behavior is influenced by people’s beliefs, many of which are based on knowledge about the behavior and its outcomes (Ajzen, 1991; Glanz & Bishop, 2010). For example, people’s understanding of skin cancer risk and facts regarding the benefits of using sunscreen correlates with more sunscreen use (Grubbs & Tabano, 2000), and parental knowledge about vaccines correlates with vaccine uptake in young children (Smith et al., 2017). Interventions with more behavioral recommendations may thus provide unique information that is applicable to the target behaviors and exert an influence through changes in information.

Multiple-behavior interventions may also affect the motivation to enact specific behaviors, another key determinant and mediator of behavioral change that comprises attitudes, social norms, and intentions (Ajzen, 1991; Rich et al., 2015; Topa & Moriano, 2010). In a meta-analysis of condom use interventions, attitudes and beliefs predicted intentions to use condoms in future encounters (Albarracín et al., 2005). Furthermore, prior work has noted the potential for motivation to spill over between health behaviors. For example, engagement in physical activity predicts better regulation of eating behaviors, and this relation is mediated, in part, by increased intrinsic motivation to exercise (Mata et al., 2009). Similarly, exercise is associated with greater motivation to quit smoking, which, in turn, is associated with higher odds of staying abstinent (LaRowe et al., 2022).
The number of behavioral recommendations may also promote behavioral change by improving behavioral skills. Behavioral skills, people’s ability to implement routines that promote successful performance of specific behaviors, are key to ensuring behavioral change. For example, interventions targeting HIV prevention, treatment adherence, and weight loss have found that skills in monitoring one’s own behavior and making corrections if goals are not being met are both important to changing behavior (Albarracín et al., 2005; Burke et al., 2011; Venditti et al., 2014). Furthermore, a recent meta-review of the health-behavior-change literature identified skills-training as one of the intervention features most often associated with reduced engagement in risk-taking behaviors, such as smoking, unhealthy eating, and alcohol use (Protogerou et al., 2020). Recommendations to change multiple behaviors may thus facilitate the acquisition and development of behavioral skills.

All in all, we expected changes in information, motivation, and behavioral skills to reflect behavior-specific processes that require attention to each behavior and may thus pose limits in terms of the number of behaviors that are of benefit. We recognize that knowledge about one behavior may be relevant to other behaviors, changing attitudes toward one behavior can produce unintended changes in other attitudes (Fishbach et al., 2004; Glanz & Bishop, 2010; McDonald et al., 2017; Pintrich, 2000; Stedry & Kay, 1966), and some skills cut across domains. Nevertheless, reviewing information, behavioral outcomes, and skills necessary to quit smoking is sufficiently time consuming to prevent also focusing on information, behavioral outcomes, and skills about using condoms. Therefore, even if information, motivation, and behavioral skills could generalize over time, the process is likely to be more demanding than changes in one behavior cuing changes in other behaviors or improvements in psychological well-being reducing the need to manage stress by smoking or having risky sex.
The Present Meta-Analysis

We conducted a comprehensive meta-analysis of multiple-behavior interventions to assess the effects of the number of recommendations included in an intervention on behavioral and clinical improvement. Interventions spanned the domains of physical exercise, diet, smoking, HIV prevention, HIV testing, HIV treatment, alcohol use, and substance use. In total, we synthesized changes in 58 different behavioral and clinical outcomes into effect sizes that represent behavioral improvements in exercise; diet; smoking; HIV prevention, testing, and/or treatment; alcohol use; or substance use. In addition, we synthesized changes in measures that represent changes in psychological well-being (e.g., depression and anxiety), health information (e.g., exercise and diet), motivation (e.g., perceived importance and attitude), and behavioral skills (e.g., stress management and decision-making skills), all of which were considered possible mechanisms of change. Although the effect sizes could also be calculated by comparing treatment groups to control groups, a problem associated with this between-group comparison approach is that different trials may employ different types of control groups (i.e., passive vs. active) and systematically affect the computed effect sizes (Karlsson & Bergmark, 2015). Specifically, a given trial with a passive control is likely to result in larger between-group comparison effect size computations than another trial with an active control, which can be regarded as receiving a minimum amount of treatment (see, Brookmeyer et al., 2016).

The main analyses took advantage of recently developed meta-regression models of correlated and hierarchical effects with three levels to account for non-independent observation from both data hierarchy (i.e., measures nested under trial conditions/samples that were nested under trials) and correlated sampling errors (e.g., multiple measures from a given trial condition). One focus of the analyses was to examine the strength and linearity of the effect of the number of...
behavioral recommendations on behavioral and clinical change. As the number of recommendations has generally accrued naturalistically in the literature (Cochran & Chambers, 1965), we used propensity-score modeling as an optimal way of controlling for differences among conditions with different numbers of recommendations and improving confidence in our inferences (Austin, 2011).

To explicate the effect associated with the number of behavioral recommendations, we tested hypotheses involving both general and behavior-specific processes, including (a) changes in some outcome behaviors cuing changes in another outcome behavior and (b) changes in the mediators of psychological well-being, information, motivation, and behavioral skills. The behavioral cuing effect and changes in psychological well-being may coexist with linear changes in the number of recommendations, as these processes do not require substantial cognitive resources to function. In contrast, mediation through information, motivation, and/or behavioral skills may be associated with a curvilinear relation between the number of recommendations and behavioral and clinical change, as these processes require greater cognitive resources to take effect. The meta-analysis was preregistered at https://osf.io/mq5kr/?view_only=54d04d57b5ef45aaa480a47259cb91cd. For a list of deviations from the preregistration and coverage of preregistered analyses not included in this article, see Supplement A.

Method

Review and Inclusion Criteria

Eight different behavioral research domains were pre-identified to be included in the search: (a) exercise, (b) diet, (c) smoking, (d) HIV prevention, (e) HIV testing, (f) HIV treatment, (g) alcohol use, and (h) other substance use interventions. We conducted searches
using MEDLINE, PsycINFO, EBSCO, Scopus, the Web of Science Core Collection, JSTOR, and Crossref. In addition to these database searches, we conducted searches of conference abstracts and ProQuest for unpublished literature. When reports were in languages other than English, they were translated and screened.

**Behavioral Domain Searches**

The following keywords were used to find intervention reports: “intervention,” “health education,” “persuasion,” “recommendation,” “treatment,” “educational program,” “rehabilitation,” “counseling outcomes,” “treatment outcomes,” “treatment effectiveness evaluation,” “treatment compliance,” “health promotion,” “behavioral change,” and “randomized trial.” Within each of the eight domains (exercise, diet, smoking, HIV prevention, HIV testing, HIV treatment, alcohol use, and substance use), these general keywords were combined with unique keywords to identify domain-specific intervention reports, as explained below.

1. To identify **diet interventions**, we used the following keywords: “binge eating,” “body image,” “bulimia,” “caloric intake,” “craving,” “diet,” “dietary restraint,” “eating behavior,” “eating disorders,” “fat intake,” “food intake,” “fruit intake,” “metabolism disorders,” “healthy nutrition,” “obesity,” “sugar intake,” “vegetable intake,” “weight control,” “weight loss,” and “healthy eating.”

2. To identify **exercise interventions**, we used the following keywords: “aerobic exercise,” “body image,” “physical activity,” “sport training,” “strength training,” “weight control,” “weight loss,” “lack of exercise,” “walking,” “gymnastics,” “going to gym,” “running,” “biking,” “work out,” and “physical inactivity.”

3. To identify **smoking interventions**, we used the keywords (a) “tobacco” and (b) “smoking.”
4. To identify **HIV prevention, HIV testing, and HIV treatment interventions**, we used the keywords “HIV,” “AIDS,” “STI/STD,” “condom use,” “circumcision,” “alcohol use,” “drug use,” and “adherence.”

5. Finally, to identify **alcohol use and substance use interventions**, we used the keywords (a) “alcohol,” (b) “drug,” and (c) “substance.”

The specific database search queries are shown in Supplement B. The literature search yielded a total of 42,065 studies with the cut off being July 27th, 2022.

**Inclusion Criteria**

Figure 1 presents a graphic depiction of the search and selection strategy. Reviewers first completed multiple rounds of training in order to establish high reliability in screening records for eligibility. Once the training and searches were complete, each reviewer was assigned a subset of records to independently screen, with each record screened by one reviewer. All records that were initially selected for inclusion and coded were later rechecked by a second reviewer to ensure the inclusion criteria were correctly applied. We applied the following inclusion criteria (Fleiss’ Kappa = 0.70 for inclusion/exclusion):

1. **Presence of at least two groups.** Reports had to include both a multiple-behavior intervention group and a comparison group that we could use as a reference. Comparison groups were separated into three categories: (a) passive controls (no recommendations); (b) single-behavior interventions (i.e., interventions making one recommendation); and (c) other multiple-behavior interventions (i.e., interventions making two or more recommendations).

2. **Presence of two or more behavioral recommendations in the interventions and account of behavioral recommendations.** Reports had to include a description of the
target intervention with enough details to verify whether multiple recommendations were included in at least one study condition. As an example, in an included report, Amaro et al. (2007) recommended that participants both reduce their substance use and reduce their sexual risk through abstinence. We excluded reports that did not provide a clear enough description of interventions to determine the number of recommendations (5%).

3. **Presence of behavioral recommendations from two or more domains.** Reports had to include an intervention that included recommendations in at least two behavioral domains which included physical activity, diet, smoking, HIV prevention, HIV testing, HIV treatment, alcohol use, and drug use. For example, Kalichman et al. (2014) included behavioral recommendations to reduce alcohol consumption and increase condom use (i.e., one alcohol use recommendation and one HIV prevention recommendation). As another example, Dushay et al. (2001) recommended participants to test for HIV (one testing behavior) and use condoms (one preventive behavior). In contrast, an intervention that recommended people to reduce their fat intake and reduce their sugar intake would not be included because it would have concerned only the dietary domain.

4. **Presence of at least one behavioral or clinical outcome.** Reports had to include at least one behavioral outcome or one clinical outcome. We excluded reports that included only information on behavioral intentions, attitudes, social norms, or other non-behavioral or non-clinical outcomes.

5. **Presence of appropriate statistics to estimate improvement over time.** Reports had to include enough statistical information to calculate effect sizes representing improvement or worsening over time. Thus, reports had to include outcome values at both the baseline and at least one posttest. When information was reported only for a delayed posttest,
reports were excluded (1%). Studies that reported only analyses with statistical controls or other statistical corrections were also excluded.

After applying our inclusion criteria and excluding studies that appeared to meet the criteria, such as secondary reports (e.g., Hershow et al., 2020) or those focusing on merely one health domain (e.g., Cunningham et al., 2020), 364 studies contributing 803 groups and 6,819 effect sizes were included in our analyses (see Supplement C for study group characteristics and Supplement D for citations of included studies). The groups from a given study typically represent different intervention conditions. However, when reports stratified results by demographics or risk factors (e.g., high- vs. low-risk drinkers; Doumas et al., 2017), we separated effect sizes for different groups.

**Coding of Effect Sizes**

All coding and data extraction were conducted via an internally produced online data entry system using a coding and entry manual developed by the research team. One researcher extracted the means and standard deviations or proportions for each available study outcome. A second researcher then rechecked this data entry for each report.

Our effect size of interest was within-group improvement over time ($d$) for all intervention-relevant behaviors and clinical outcomes: If a study reported multiple outcomes, then multiple effect sizes were calculated. If there were multiple posttests, then multiple delayed effect sizes were calculated. To calculate $d$ for continuous outcomes, the mean of the pretest measure was subtracted from the mean of the posttest measure, and this difference was then divided by the pooled standard deviation of the means (Borenstein et al., 2009). To calculate $d$ for proportional outcomes, an odds ratio was calculated by dividing the odds of the behavior at the posttest by the odds of the behavior at the pretest and then converting it into a $d$ by taking the
natural log and dividing it by 1.81 (Borenstein et al., 2009). In cases where the reported proportions were 0s or 1s, 0.005 was added to 0s or subtracted from 1s, respectively, to reduce extreme odds ratio calculations (Sweeting et al., 2004). All effect sizes were corrected for small sample size bias (Hedges & Olkin, 1985) and coded such that positive effect sizes reflect health improvement (e.g., increased testing rates and decreases in unprotected sex). As the correlation between pretest and posttest observations was unknown, we assumed that \( r = 0.5 \) in calculations (Morris, 2008). We used corrected estimates and standard errors for studies reporting results from three reports of cluster-randomized trials (Hedges, 2007). Previous analyses from other domains of the larger dataset have found that models are robust to changes in the assumed correlation (Wilson et al., 2015).

We applied these calculation rules to obtain effect sizes for multiple posttests as well. If a study had three measurements, such as a baseline, an immediate posttest, and a delayed posttest, we calculated two effect sizes: a pretest vs. immediate posttest effect size and a pretest vs. delayed posttest effect size. We used pretests as the comparison throughout, as change relative to the baseline reflects improvement at each point. Each effect size has an associated time variable indicating whether the comparison group is from an immediate posttest or a delayed posttest as well as a measure of how many days post-intervention the measurement was obtained.

The dataset included 58 different outcomes, which were categorized as being either behavioral or clinical. Behavioral outcomes are executed by individuals and are observable through direct measurements, whereas clinical outcomes require biomedical samples that are analyzed and tested in a laboratory setting and do not reflect behavior in a direct way. Examples of behavioral outcomes included attending counseling sessions, using condoms, adhering to a
medication regimen or a diet, performing physical activity, and smoking. Examples of clinical outcomes were blood glucose level, drug urine analysis, HIV viral load, and death.

The outcomes could also be categorized according to the domains (see Supplement E). Relevant to lifestyle interventions, in the domain of diet, commonly measured outcomes were energy intake (e.g., kcal/week); the intake of specific nutrition components (e.g., carbohydrate, protein, vitamin, water); number of meals per day; compliance with dietary recommendations (e.g., fruit and vegetable intake); and presence of unhealthy eating or overeating. Commonly measured outcomes related to physical activity were the presence of daily exercise behaviors, total hours of weekly physical activity, presence of occupational physical activity, compliance with exercise recommendations, sedentary behaviors, self-monitoring of pulse and blood pressure, energy expenditure in physical activity, and number of hours watching TV. Commonly measured outcomes related to smoking were whether one was smoking and the number of cigarettes smoked per day.

In the area of HIV interventions, commonly measured outcomes related to sexual risk were whether participants had engaged in risky sexual behaviors, such as not using condoms during sexual intercourse or having multiple sexual partners. Commonly measured outcomes related to testing were whether one was tested for HIV or sexually transmitted infections (STIs), and commonly measured outcomes related to treatment were whether one received treatment for HIV or STIs, problems adhering to medications, and changes in biomedical indices (e.g., viral load) for receiving antiretroviral therapy.

In the area of alcohol and substance use interventions, commonly measured outcomes related to alcohol use were the amount of alcohol consumption during a temporal unit (e.g., day); the frequency of excessive alcohol use (e.g., number of excessive drinking days in a month); the
number of days being drunk; and the proportion of abstinence days. Commonly measured outcomes related to *substance use* were the amount of substance use (e.g., heroin, opioids, methadone, cocaine, crack) during a temporal unit (e.g., in the past week); the frequency of excessive drug use; the frequency of ecstasy use; the proportion of abstinence days; and the number of times one had sexual intercourse under the influence of drugs.

We calculated change in all the behavioral and clinical outcomes measured in a study (see Supplement F for sample reports). For example, if a lifestyle intervention recommended exercise and diet but also measured smoking, all three behavioral changes were recorded. In addition, the dataset included effect sizes for improvements in different measures corresponding to psychological well-being. Thus, if an intervention condition reported a measure of depression and a measure of coping behavior, which are two psychological well-being measures, the average of the two effect sizes was used to gauge changes in psychological well-being, which was used as a mediator.

**Psychological well-being.** We obtained effect sizes for the following measures related to well-being: (a) depression and mania, reverse-scored through our effect size calculation (Calgary Depression Scale; Addington et al., 1993; Center for Epidemiological Studies-Depression; Radloff, 1977; Patient Health Questionnaire-9; Kroenke et al., 2001; Clinical Global Impressions Scale-Bipolar Version-Overall and Mania + Depression subscales; Spearing et al., 1997; Montgomery–Åsberg Depression Rating Scale; Montgomery & Asberg, 1979; Edinburgh Postnatal Depression Scale; Cox et al., 1987; Hamilton Depression Rating Scale; Hamilton, 1960; Beck Depression Inventory; Beck et al., 1961; mental health subscale of the Addiction Severity Index; McLellan et al., 1992); (b) psychological quality of life (Medical Outcomes Study 36-item short-form health survey; McHorney et al., 1992; Health Promoting Lifestyle
scale (Walker & Hill-Polerecky, 1996); EuroQol Visual Analogue Scale; EuroQol Group, 1990; Acceptance and Action Questionnaire–II; Bond et al., 2011; self-reported mindfulness; self-reported trauma); (c) anxiety and stress, reverse-scored through our effect size calculation (Symptom Checklist-90; Derogatis et al., 1973; anxiety subscale of the Depression Anxiety Stress Scale-21; Henry & Crawford, 2005; self-report items); (d) social quality of life (Social Functioning item from the Medical Outcomes Study 36-item short-form health survey; Ware & Sherbourne, 1992; community engagement, social support, and social function items; self-reported victimization); (e) positive mood (Positive and Negative Affect Schedule; Watson et al., 1988; Profile of Mood States; McNair et al., 1971); and (f) coping (high self-reported positive reinterpretation; low self-reported behavioral disengagement).

As mentioned previously, the coding of effect sizes took the form of within-group growth over time. The advantage of this approach as opposed to between-group comparisons is that different trials may employ different types of control groups (i.e., passive vs. active; see Karlsson & Bergmark, 2015) and that control and intervention groups in our meta-analysis generally have different numbers of recommendations (e.g., 0 and 3 in one study versus 1 and 2 in another). One advantage of the within-group growth approach is to allow each group to serve as its own control over time and different interventions to be compared as a function of the number of recommendations. Passive and active control groups could receive the coding of 0 and 1, respectively, regarding the number of behavioral recommendations, and different interventions could be coded in terms of the number of behavioral recommendations as well.

**Coding of the Number of Recommendations**

As was done in prior analyses of earlier, specific subsets of these data (Dai et al., 2020; Sunderrajan et al., 2021; Wilson et al., 2015), we coded the number of recommendations as the
number of specific, observable behaviors participants received for general health promotion in the intervention programs. The recommendations are behavioral goals for intervention recipients to follow. These recommendations were extracted from authors’ descriptions of the intervention methodology. For example, Go et al. (2015) encouraged participants to reduce (a) their sexual risk behavior and (b) their injection drug use. Therefore, that condition was coded as having two behavioral recommendations. In another study, Kypri and McAnally (2005) reported an intervention that included four recommendations, namely (a) increasing physical activity, (b) reducing alcohol use, (c) increasing fruit intake, and (d) increasing vegetable intake. The intercoder reliability coefficient for the coding of the number of recommendations was acceptable ($r = .81$).

**Coding of Other Moderators**

We coded a variety of moderators that can be used to describe the sample and analyze possible covariates. The interrater reliability for each appears in Table 1.

**Report characteristics.** We coded (a) publication year; (b) first authors' institution (e.g., college, hospital, research center); (c) first authors' institutional area (e.g., psychology, public health, medicine); (d) source type (e.g., journal article, dissertation); (e) language of the intervention; (f) country of the intervention (as well as state or city if provided); and (g) research design.

**Sample features.** We recorded the following sample characteristics: (a) the percentage of males in each group; (b) the percentage of male participants who identified as gay or bisexual; (c) the percentage of participants of European, African, Latin, Asian, and Native American descent; (d) the mean age; (e) the country or countries where the study was conducted; and (f)
the proportion of the sample with a pre-existing health condition, defined as a previous diagnosis of a relevant disease (e.g., diabetes in lifestyle domain studies).

**Method features.** For each study, we recorded factors related to the design and implementation of the intervention, including (a) whether participants were randomly assigned to the intervention and control groups; (b) the mean number of days between the intervention and each posttest; (c) whether participants were recruited in a clinical or non-clinical setting; (d) whether an expert source/facilitator was present; and (e) whether the intervention was delivered to a group, to individuals, or to a combination of the two. We also recorded (f) whether the primary exposure format was face to face or another format such as video; (g) whether the exposure setting was clinical (e.g., a clinic) or non-clinical (e.g., a school); (h) whether the study used intention-to-treat analyses; (i) the total time participants spent participating in the intervention; and (j) whether the authors described the intervention as culturally appropriate, tailored, or neither.

**Analyses**

**General approach.** All the analyses were conducted in R with the alpha level set at .05. The primary unit of analysis was the effect size $d$ representing behavioral and clinical improvements over time and weighted by its precision, which is the inverse of the variance (Borenstein et al., 2009). Meta-regression was adopted as the primary analytical approach. Given that the present data involved 6,819 effect sizes from 803 independent samples/groups from 364 studies and that each independent sample was associated with more than eight effect sizes on average, our primary analysis involved the model of correlated and hierarchical effects with three levels (CHE +; Pustejovsky & Tipton, 2021) using the R packages Metafor (Viechtbauer, 2019) and ClubSandwich (Pustejovsky, 2021) with other analyses (i.e., publication bias, propensity-
score modeling, and mediation testing) performed using the packages Robumeta (Somaa et al., 2021), ordinal (Christensen, 2011), Psych (Revelle, 2015), and Lavaan (Rosseel, 2012).

The CHE+ model was recently developed to merge the hierarchical approach with the robust variance estimation (RVE) approach in meta-regression. This model assumes a more generalized form than both hierarchical and RVE approaches by accounting for effect size dependence arising from both a nested data structure (e.g., measures nested under trial conditions/samples that are themselves nested under trials) and correlated error variances among the effect sizes from a given sample (e.g., sampling error). In addition, unlike the RVE approach, which relies on moment-of-methods estimation, the CHE+ model uses the restricted maximum likelihood (REML) estimation to obtain parameter estimates. Both empirical and simulated analyses have suggested that the CHE+ models can reach more precise (i.e., smaller standard errors) and accurate (i.e., estimates being closer to the true values) parameter estimates than RVE models (Pustejovsky & Tipton, 2021). All these features made the CHE+ model a good option for the current research.

In specifying a given CHE+ model, in addition to the fixed-effects model part, we included three random-effects parameters at default to account for random variability of the true effects at the level of study (level 1), sample (level 2), and measure (level 3). We also specified the correlation parameter of sampling error (i.e., $\rho = 0.6$) with smoothing of the marginal variances in the RVE-like part of the CHE+ model and used small-sample corrections (McCaffrey & Bell, 2003; Pustejovsky & Tipton, 2018). These general strategies were adjusted according to specific research questions in analyses, which included (a) publication bias determination, (b) covariate identification, (c) propensity score models, and (d) tests of the main hypotheses.
Publication bias analyses. To detect publication bias or selective reporting in the dataset, we applied multiple methods, including (a) the funnel plot for direct visualization (Light & Pillemer, 1984); (b) the Egger’s Sandwich test to quantify the asymmetry by detecting whether effect sizes are correlated with sample size (Pustejovsky & Rodgers, 2019); and (c) the selection models to determine whether effect sizes were relatively over-represented in certain range(s) of \( p \)-values (Vevea & Hedges, 1995). In the funnel plot, all effect sizes were aggregated to the sample level prior to generating the plots. In the Egger Sandwich test, the effect sizes were not aggregated, and statistical dependence was handled with both the RVE (with \( \rho = 0.6 \)) approach and the hierarchical approach due to a lack of past research with the CHE+ version of Egger’s Sandwich test. In these tests, we used the variance term in predicting effect sizes (Rodgers & Pustejovsky, 2021) and calculated a modified version of the variance according to Pustejovsky and Rodgers (2019) to mitigate artifactual association with effect sizes. The selection model relied on data aggregated to the sample level and examined the relative probability density of observing one-sided \( p \)-values using the following \( p \)-value cutoff points: .001, .025, .05, .1, .5, and 1. Given that past research has shown the predictive power of the number of behavioral recommendations on effect sizes, we also included the number of behavioral recommendations as a covariate in the fixed-effect part to improve the fit of the selection model.

Controlling for important covariates. To control for known covariates that may predict behavioral and clinical outcomes, we conducted CHE+ models to analyze univariate associations between our effect sizes and the following exploratory moderators: (a) whether the effect size was based on clinical or behavioral outcomes, (b) whether the exposure to the intervention was as an individual or as part of a group, (c) whether the intervention was delivered in a clinic or
not, (d) whether the intervention facilitator was a professionally trained expert, (e) whether the intervention involved motivational strategies, (f) whether the effect size was calculated using proportional outcomes (and later transformed into Cohen’s \( d \) metric), (g) a socio-economic status variable indicating the proportion of participants completing high school education, (h) the standardized duration of the intervention program, and (i) the standardized time interval between baseline and posttest measurement. Most of these covariates were suggested by previous meta-analyses synthesizing domain-specific evidence (cf. Dai et al., 2020; Sunderrajan et al., 2021; Wilson et al., 2015) and, if they predicted effect sizes, were included in the CHE+ models that tested how the number of behavioral recommendations affected behavioral and clinical change.

**Propensity score model.** To further improve inferences about the effects of the number of behavioral recommendations while controlling for possible covariate differences among studies with different recommendation numbers, the CHE+ models included a scalar version of the propensity score as another covariate (Joffe & Rosenbaum, 1999). The propensity scores were estimated from a generalized linear mixed-effects model, wherein linear combinations of available covariates were introduced as predictors of the number of behavioral recommendations via a logistic link function. The number of behavioral recommendations was treated as an ordinal outcome including 0 recommendations, 1 recommendation, 2 recommendations, 3 recommendations, and 4 or more recommendations.\(^1\) The propensity score model was based on 803 condition-level observations and included a random effect at the trial level. Overall, developing the propensity score model was an iterative process. The model selection decision was based on criteria that included (a) the degree of balancing covariates among the number of

\(^1\) The propensity score model represents the only occasion where the number of recommendations was categorized for the sake of modeling. In all other models, the number of recommendations was treated as a continuous variable.
behavioral recommendations on each quintile of the propensity score and (b) the area of common support regarding the propensity score distribution. Also, preference was given to (c) models with a relatively simple linear combination of covariates (rather than interactions and quadratics; see Joffe & Rosenbaum, 1999) and (d) models preserving observations during the trimming process. The procedures generally followed the recommendations outlined in Bergastra et al. (2019).

**Analyses testing main hypotheses.** Using data associated with the untrimmed trial conditions from the propensity score model and those from all the available observations, we implemented the CHE+ model to examine associations between the number of behavioral recommendations and behavioral and clinical improvements. We hypothesized that tapping a general process could result in a monotonic association between the number of behavioral recommendations and behavioral and clinical change, mathematically expressed by a linear term in the CHE+ model. In contrast, the number of behavioral recommendations may predict improvements only up to a point, approximating an inverted-U shape mathematically expressed by a quadratic term in the CHE+ model. To avoid collinearity between the linear and quadratic terms of the number of behavioral recommendations, we centered the linear term before squaring it to generate the quadratic term.

We then implemented two CHE+ models. The first model included the linear term and all the covariates, and the second included the linear and quadratic terms along with the same set of covariates. As a reminder, the covariates included an indicator variable for the type of effect size (i.e., clinical outcome vs. behavioral outcome) in the model; the covariates associated with intervention features; a modified version of effect size variance; and the propensity score estimates. If clinical outcomes were shown to be different from behavioral ones, we would then
explore the relation with the number of behavioral recommendations separately for each outcome.

**Evaluation and sources of heterogeneity.** To gauge the heterogeneity of our effects, we obtained $Q$, $I^2$, and 95% prediction interval (PI) of the base model, which is a CHE+ model with only the intercept term in the fixed-effects part. We computed $I^2$ by dividing the estimated study-level variance (i.e., $\sigma_1^2$) by the total variance estimated from all three levels of data hierarchy and computed 95% PIs using the estimated average effects and the study-level variance in the base model. $I^2$ is a statistic that indicates the proportion of heterogeneity, whereas the 95% PI is an absolute description of effect size dispersion (Borenstein et al., 2017). When substantial heterogeneity was identified, we performed exploratory analyses to investigate sources of heterogeneity.

**Analyses of our hypotheses.** We began to test mechanisms by examining cuing, which assumes that more behavioral recommendations can potentiate each other when one behavior cues other behaviors. For these analyses, we aggregated effect sizes to the sample level and created the categories of (a) exercise, (b) diet, (c) smoking, (d) HIV (including prevention, testing, and treatment), (e) alcohol use, and (f) drug use. Because the literatures tend to intervene on exercise, diet, and smoking together (e.g., an intervention to reduce cardiovascular disease), and similarly on HIV, alcohol use, and drug use together, these behaviors present two literature groups, including lifestyle (subsuming exercise, diet, and smoking) as well as HIV, alcohol, and substance use. We tested the cuing effect among behaviors within the lifestyle literature and within the HIV, alcohol, and substance use literatures. In these analyses, specific behaviors measured at the immediate posttest were set to predict other behaviors measured at the delayed posttest, while controlling for potential selection bias and the interval of time between the
baseline and posttest measurements. Because each sample generated only one observation, the
cuing-testing models included two random effects that corresponded to the study and sample
levels, respectively. For instance, in the model testing whether exercise would cue diet and
smoking, changes associated with exercise at the immediate posttest as well as time interval were
included in the 2-level meta-regression model to predict changes in diet and smoking at the
delayed posttests.

We also tested whether the effects of the number of recommendations on behavioral and
clinical changes were mediated by changes in psychological well-being, information, motivation,
and behavioral skills. We followed the 3-step tradition in testing an “X-M-Y” mediational chain.
The steps included testing (a) X-Y (i.e., the number of behavioral recommendations predicting
effect sizes); (b) X-M (i.e., the number of behavioral recommendations predicting the mediator);
and (c) XM-Y (i.e., both the number of behavioral recommendations and the mediator predicting
behavioral and clinical improvements). We focused on the product of the recommendation-
mediator path coefficient in Step X-M and the mediator-behavioral/clinical change path
coefficient in Step XM-Y because such a product indicates mediation strength (i.e., indirect
effect) without involving the measurement unit of the mediator (Hayes & Rockwood, 2017).
Three CHE+ models were fit to immediate and delayed changes and each CHE+ model’s fixed-
effect part corresponding to a given step of the mediational chain with a covariate indicating the
elapse of time (in days) between the intervention and the dependent measure. When a given
mediator showed a meaningful magnitude of mediation strength (i.e., indirect effect), we further
conducted tests of corresponding indirect effect(s) using Sobel tests (Sobel, 1982) and path
analyses with both bootstrapped confidence intervals from 5000 iterations (Shrout & Bolger,
2002) and Huber-White robust standard error (Huber, 1967; White, 1980).
Transparency and Openness

The pre-registration of the review’s protocol is at

https://osf.io/mq5kr/?view_only=54d04d57b5ef45aaa480a47259cb91cd. We prepared the
protocol by following the PRISMA-P checklist and reported our work by following the PRISMA
guidelines (see Supplement L). The R code and data supporting the findings reported in this
study are openly available in OSF at

https://osf.io/qsx57/?view_only=6fffd7080d844e98cbd2d0834ed61e9. The present paper
includes deviations from the pre-registration. The major deviation involves using the more recent
and more generalized meta-regression models of correlated and hierarchical effects (CHE)
instead of the pre-registered robust variance estimation (RVE). In addition, a couple of questions
focusing on the refined effect of specific recommendation combinations, sample characteristics,
methodological factors on intervention efficacy could not be fully tested and reported due to the
constraints arising from both data availability and the scope of the present paper. Details about
specific deviations are described in Appendix A in the supplement.

Results

Description of Included Studies, Samples, and Effect Sizes

Table 1 presents a description of our data. Overall, we included a total of 6,878 effect
sizes (with positive values indicating improvement in the direction of healthy behavior and
clinical outcomes) from 803 independent groups from 364 studies that sampled 186,729
participants. The effect sizes included 4,282 immediate post-intervention effect sizes from 781
independent groups from 355 studies with an average trial-onset-to-measurement time of 215
($SD = 292$) days and 2,596 delayed post-intervention effect sizes from 378 independent groups
from 171 studies with an average trial-onset-to-measurement time of 385 ($SD = 267$) days. The
large majority (96%) involved random group assignment. The median group contributed 5 effect sizes, and the average group contributed 8.6 effect sizes. The average number of recommendations across all groups was 2.50, and the average number of recommendations among multiple-behavior interventions was 3.37. All the studies were reported in English except for one study reported in Korean and another study reported in Italian. Nearly all (99%) were from journal articles having a median publication year of 2016 (range: [1979, 2022]), with the remainder coming from doctoral dissertations. Sample sizes varied widely ($M_{\text{size}} = 513, SD_{\text{size}} = 1,295$), and the average sample was racially/ethnically diverse, with an average racial/ethnic background of 46 percent Non-Hispanic White, 27 percent African, 16 percent Asian, 13 percent Latinx, 1 percent Native American, and 14 percent Other or Multiple. The average sample age was 40 ($SD = 16$) years old and had good representation of males and females (57% female, $SD = 31$%), with 45 percent completing high school or an equivalent. Among the 364 included studies, most were conducted in the United States (51.4%), with much smaller proportions elsewhere: United Kingdom (5.5%), The Netherlands (3.6%), Australia (3%), Brazil (2.5%), Canada (2.5%), China (2.5%), Japan (2.2%), Germany (1.9%), India (1.9%), and Spain (1.9%). Supplement G lists details of study distribution over countries and how each country is categorized into Western vs. non-Western countries (World Population Review, 2023).

**Assessment of Bias**

Figure 2 displays funnel plots regarding the effect size distribution according to the precision measure of standard error after aggregating effect sizes within each sample. A visual inspection suggests no clear asymmetry of effect size distribution. Beyond the funnel plots, results from the RVE and hierarchical variants of the Egger’s Sandwich test suggested no
evidence of effect size distribution asymmetry in either the hierarchical version \((B = -0.26, SE = 0.13, p = .98)\) or the RVE version \((B = -0.09, SE = 0.11, p = .78)\) of the test.

In addition, results from the selection model showed that, compared to the unadjusted model, the adjusted model fit significantly better \((p < .001)\) by including the relative probability density weights associated with the \(p\)-value intervals with higher weights suggesting a stronger tendency to report findings for particular ranges of \(p\)-values. Specifically, the following weights (which are all significant at \(p < .001\)) were estimated:\n\(B = 2.56\) for \(p\) between \(.001\) and \(.025\); \(B = 2.57\) for \(p\) between \(.025\) and \(.05\); \(B = 2.23\) for \(p\) between \(.05\) and \(.1\); \(B = 3.65\) for \(p\) between \(.1\) and \(.5\); and \(B = 0.98\) for \(p\) between \(.5\) and \(1\). Because the first \(p\)-value interval (i.e., from \(0\) to \(.001\)) was used as a reference by constraining it to have a weight of \(1\), this pattern of results did not suggest a clear bias towards including more effect sizes associated with relatively smaller \(p\)-values, as the weights are generally balanced among all categories. In sum, the assessments did not evidence bias in the present data, although we included the modified variance (Pustejovsky & Rodgers, 2019) as a covariate in all the subsequent meta-regressions to enhance our estimates of the moderators of interest.

**Tests to Identify Important Covariates**

Results from the preliminary CHE+ models suggested that clinical outcomes were associated with smaller effect sizes than behavioral outcomes \((B = -0.09, SE = 0.03, p = .004)\). In addition, among important covariates related to method features, individual or group exposure to interventions \((B = 0.02, SE = 0.03, p = .36)\) and interventions including an expert facilitator \((B = 0.03, SE = 0.03, p = .30)\) did not significantly moderate behavioral and clinical improvement. The socio-economic variable indicating the proportion of participants completing high school education \((B = 7e-4, SE = 6e-4, p = .24)\) did not significantly predict behavioral and clinical
improvement either. In contrast, interventions that took place in a clinical (vs. non-clinical) setting ($B = 0.07, SE = 0.03, p = .02$), interventions that used motivational strategies ($B = 0.30, SE = 0.04, p < .001$), longer intervention durations ($B = 0.06, SE = 0.02, p = .01$), a shorter time interval between baseline and posttest ($B = -0.03, SE = 0.01, p = .005$), and proportional outcomes ($B = 0.27, SE = 0.08, p < .001$) were associated with greater behavioral and clinical improvements. As a result, we included these significant confounds in the CHE+ models that tested the impact and processes of the number of behavioral recommendations. Yet, these covariates could not be included in the analyses of cuing and mediational processes because they led to very small numbers of observations (i.e., $k < 70$; for recommendations about minimum sample sizes, see Tanner-Smith & Tipton, 2014).

**Estimates of Propensity Scores**

The selection of variables for the propensity score model was based on the criterion that the variables must be considered as “baseline covariates” and cannot be affected by the number of behavioral recommendations, which is the “treatment.” Checking covariate balance among five categories of the number of recommendations required dividing the trial conditions into quintiles based on the propensity score, resulting in $5 \times 5 = 25$ combinations. Fifteen out of 132 categorical covariates were selected given low missing rates (i.e., < 8%) and sufficient observations (i.e., > 100) in each category of the covariate. Table 2 shows more details about the covariates. The final propensity score model consisted of regressing the number of behavioral recommendations on the main effects of these covariates with a random effect at the trial level. Figure 3 illustrates the contrast prior to and after trimming the trial conditions given the estimated propensity scores. Further details regarding the covariate balance conditional on propensity score quintiles can be found in Supplement H. Overall, 203 trial conditions were
trimmed off in identifying the area of common support with the propensity score ranging from \([-6, 0]\). This process reduced the total number of trial conditions from 803 to 600, leading to 120 trial conditions in each final quintile of the propensity score.

**Effect of the Number of Recommendations in an Intervention**

Given potential differences between behavioral and clinical effect sizes, we first tested the linear vs. nonlinear hypothesis separately for behavioral and clinical improvements. Table 3 shows the CHE+ model results, indicating a positive, linear effect but not a quadratic effect ($p > .13$) of the number of behavioral recommendations on both behavioral and clinical improvements. Table 4 shows the main CHE+ model results, indicating that the linear form ($B = 0.03, SE = 0.01, p < .001$) but not the quadratic form ($B = -2e-3, SE = 1e-3, p = .12$) of the number of behavioral recommendations predicted effect sizes after controlling for the confounds. Figure 4 illustrates the positive linear effect of the number of behavioral recommendations on behavioral and clinical improvements. In addition, the CHE+ models that did not include the propensity scores as a covariate (and thus used 781 trial conditions) generated a larger estimate of the positive, linear effect of the number of behavioral recommendations ($B = 0.04, SE = 0.007, p < .001$), suggesting that the propensity scores were able to attenuate an upward bias in such an estimation. The parallel effects of the number of behavioral recommendations on behavioral and clinical outcomes supported treating the two types of outcomes together in subsequent analyses. This strategy was also desirable to increase the sample size available for each analysis.

Finally, we performed three sensitivity analyses. The first sensitivity analysis followed the difference-in-differences approach by using between-group effect sizes. Results consistently supported the linear ($B = 0.03, SE = 0.01, p = .001$) but not the quadratic term ($B = -0.002, SE = 0.001, p = .12$) of the number of behavioral recommendations (see Supplement I). The second
sensitivity analysis gauged whether the present findings are robust against the concern that psychology findings are overly reliant on WEIRD (Western, Educated, Industrialized, Rich, and Democratic, see Henrich et al., 2010) samples. Results suggested that the linear effect of the number of recommendations was significant after including an indicator variable for whether participants resided in a Western vs. non-Western country as well as a socio-economic status variable indicating the proportion of participants completing high school education (see Supplement J). The last sensitivity analysis helped test the robustness of findings in Table 4 against three indicator variables related to study quality, including (a) whether the study randomized participants/clusters into conditions, (b) whether the study included a posttest at least six months after the baseline measurement, and (c) whether the study involved an intent-to-treat analysis. All results showed consistent findings regarding the linear effect of the number of recommendations (see Supplement K).

### Evaluation and Sources of Heterogeneity

We evaluated effect-size heterogeneity of the model in Table 4. Results supported the presence of a small to moderate level of heterogeneity, $\tau = 0.30$, $Q(2742) = 70770.62$, $p < .001$, $I^2 = 28.59\%$. The 95% PI [-0.29, 0.88] showed a wide range of effects in the literature we synthesized.

To explore the sources of heterogeneity, we identified seven additional study-level predictor variables in addition to the number of recommendations, three related to the study population and four related to the study design. Table 5 shows three nested models fit to the same data, including (a) an average-effect model, (b) a model that adds the number of recommendations to the average-effect model, and (c) a model adding the seven additional predictors. Table 5 also shows the percentage of reduction in heterogeneity for each model after
adding the predictor variables to the average-effect model. As shown, the number of recommendations alone accounted for 7.23% of the estimated heterogeneity. Moreover, adding the additional seven predictors to the model with number of recommendations accounted for up to 29.31% of heterogeneity. In particular, studies that included a higher number of recommendations ($B = 0.03, SE = 0.01, p = .002$), studies with longer intervention duration ($B = 0.05, SE = 0.02, p = .01$), and studies conducted in non-Western countries ($B = -0.19, SE = 0.09, p < .05$), were associated with larger effect sizes, which suggested some important sources of heterogeneity.

**Tests of Potential Mechanisms**

We tested possible behavioral cuing as well as well-being and motivation as mediators of the linear effect of the number of recommendations included in an intervention.

**Behavioral cuing.** Table 6 shows the results from the CHE+ models regarding behavioral cuing. These analyses considered, after controlling for potential selection bias and the time of measurement for both the cuing and the cued behaviors, the degree to which a specific behavior (e.g., exercise) predicts other behaviors in the study (e.g., diet and smoking), investigating whether changes in that specific behavior potentiate changes in the other behaviors. Results indicated that, after controlling for selection bias and the time of measurement, changes in HIV ($B = 0.70, SE = 0.18, p = .01$), alcohol use ($B = 0.48, SE = 0.10, p < .001$), and drug use ($B = 0.35, SE = 0.10, p = .02$) showed significant cuing effects, whereas changes in exercise ($B = 0.27, SE = 0.22, p = .26$), diet ($B = 0.14, SE = 0.06, p = .10$), and smoking ($B = 0.02, SE = 0.34, p = .97$) did not demonstrate cuing. The overall pattern suggested significant cuing across behaviors in the HIV and substance use literatures.
Changes in well-being. Along with cuing, changes in well-being, but not motivation, could explain the linear effects of the number of behavioral recommendations on behavioral and clinical change. Table 7 summarizes the results from the CHE+ model testing psychological well-being, information, motivation, and behavioral skills, respectively, as mediators of the effect of the number of behavioral recommendations on behavioral and clinical improvement. As shown, only psychological well-being and motivation showed promise, as the number of behavioral recommendations predicted improvement in psychological well-being ($B = 0.12, SE = 0.05, p = .02$) and motivation ($B = 0.07, SE = 0.03, p = .05$), and well-being ($B = 0.35, SE = 0.18, p = .107$) and motivation ($B = 0.47, SE = 0.13, p = .01$) also seemed to predict behavioral/clinical improvement when the number of recommendations was included. Conversely, in the follow-up tests of the indirect effect (i.e., product of the two paths), only well-being was evidenced as a mediator in all three tests in Table 8, including the Sobel test ($Z = 3.35, p < .001$); the path analytic model using bootstrapped CI ($B = 0.05, SE = 0.02, p < .001$); and the path analysis using robust standard error ($B = 0.05, SE = 0.02, p < .001$). Motivation was not a mediator ($ps > 0.18$). These results suggested that adding one recommendation to the number of behavioral recommendations would lead to a $d = 0.042$ increment in behavioral and clinical improvement at the posttest via improvements in psychological well-being.

Discussion

Employing the recent CHE+ models to account for the complex dependence structure among effect sizes from 364 study reports, the present meta-analysis investigated how and why the number of behavioral recommendations in an intervention affected clinical and behavioral changes in eight domains related to lifestyle, HIV, and substance use. Results indicated a positive, linear (but not curvilinear) effect of the number of recommendations on behavioral and
clinical changes after controlling for confounding factors from several sources, including selective reporting (i.e., modified variance of the effect size); conceptually important covariates (e.g., whether the effect size is associated with a behavioral or clinical measure); and 15 other covariates empirically derived and controlled for by introducing the propensity score (see Table 4 and Figure 4). These results thus offer both stronger and more generalizable causal inference about the impact of behavioral recommendations than previous work (cf. Dai et al., 2020; Meader et al., 2017; Nigg & Long, 2012; Prochaska & Prochaska, 2011; Sunderrajan et al., 2021; Wilson et al., 2015). In addition, the findings about behavioral cuing and psychological well-being point to general mechanisms that have not been previously identified and are key to designing future programs. We discuss these findings and their implications in further detail subsequently.

**The Effect of the Number of Behavioral Recommendations**

We developed a propensity score model to better isolate the impact of the number of behavioral recommendations, which could covary with other factors (Cochran & Chambers, 1965). The propensity score model allows researchers with observational data to approximately equate trial conditions, such as, in our case, groups with different numbers of behavioral recommendations (Austin, 2011). Overall, the propensity score model suggested some differences across trial conditions receiving zero, one, and higher numbers of behavioral recommendations (Figure 3A). Trimming trial conditions served to reduce but not fully eliminate the differences across the numbers of recommendations (Figure 3B). Likewise, comparing the CHE+ model parameters with and without covarying propensity scores (0.03 vs. 0.04) revealed an upward bias when propensity scores were not considered, which can result in false positive findings or exaggerated efficacy claims. Nevertheless, the propensity-score-adjusted estimate
still suggested that each behavioral recommendation contributes to positive behavioral and clinical change. For example, recommending four behaviors would result in approximately 0.10 standard deviations of clinical and behavioral improvement over not recommending any behavioral recommendations.

Figure 4 illustrates the positive linear effect of the number of recommendations on behavioral and clinical outcomes. As shown, although the number of observations is smaller as the number of recommendations exceeds four, no clear pattern of a curvilinear relation appeared in our data. This finding of a linear effect of the number of recommendations on behavioral and clinical outcomes is consistent with those of prior meta-analyses of HIV (Sunderrajan et al., 2021) and substance use (Dai et al., 2020) interventions, although one meta-analysis from the lifestyle domain suggested an inverted-U relationship (see Wilson et al., 2015). Regarding the causes of such mixed findings, despite the inclusion of different domains and study reports, the differences may be due to different analytic methods. Specifically, Wilson et al. (2015) used an ANOVA-like approach by grouping recommendation numbers into four categories (i.e., 0, 1, 2-3, 4+) and observed a small efficacy decrease from the 2-3 category to the 4+ category, whereas other meta-analyses supporting the linear pattern employed the multivariate meta-regression approach to directly test the effect of the number of recommendations. We think this linear-vs-non-linear issue will be clarified with future intervention studies including a relatively high number of recommendations beyond four. The current body of evidence supports a linear instead of a curvilinear effect of the number of recommendations on behavioral and clinical outcomes.

The linear effect of the number of recommendations can also help to understand heterogeneity in the synthesized literature. Based on the estimated $I^2$ and 95% PI, between-study heterogeneity constitutes 28.59% of the total variance, and the expected intervention effect can
range from -0.29, which is a small-to-moderate negative effect, to 0.88, which is a large positive effect. To explore the sources of this heterogeneity, we considered the number of recommendations and seven additional predictors related to study populations and methodologies. As shown in Table 5, we found that the number of recommendations alone accounted for up to 7.23% of the synthesis heterogeneity, and that higher numbers of recommendations, longer intervention duration, and involving participants from non-Western countries were associated with greater intervention efficacy. These initial findings thus suggested three methodological factors contributing to heterogeneity, which may help guide future study designs (Melsen et al., 2014).

Although the linear effect of the number of recommendations was robust to the inclusion of several research-quality factors in a sensitivity analysis (see Supplement K), other analyses may be attempted in the future. For example, although we examined whether a study used randomization, we did not capture whether that randomization used rigorous procedures (e.g., random sequence generation and allocation sequence concealment; Dechartres et al., 2016), nor did we measure other factors related to potential bias such as double blinding (Page et al., 2016) or trial pre-registration (Odutayo et al., 2017). Thus, we recommend researchers to additionally consider these important research-quality factors in future syntheses of this literature.

The General and Behavior-Specific Processes of Change

The present meta-analysis was structured around testing a framework of general processes, which do not require a substantial amount of cognitive capacity or implementation motivation, and behavior-specific processes, which do require more cognitive resources and motivation to take effect. Specifically, we proposed that, in the case of multiple-behavior interventions, general processes should lead to a positive effect of higher numbers of
recommendations because the effects of these recommendations occur efficiently. In contrast, behavior-specific processes should be associated with a curvilinear relation between the number of behavioral recommendations and behavioral and clinical change, as the increasing demand on cognitive resources should lead to diminishing returns at high numbers of recommendations.

Accordingly, we found linear associations between the number of recommendations and behavioral and clinical improvements across all tests (Tables 3 & 4). That is, there was no evidence of weakening or reversals in the positive effects of the number of recommendations when a study recommended a higher number of behaviors within the range of the present investigation. Yet, even though the number of behavioral recommendations ranged from 0 to 17, 87 percent of the conditions had four or fewer recommendations. This overall linear association thus supports the notion that general processes of behavioral change explain the impact of the number of recommendations on change in behavioral and clinical outcomes.

This linear relation also challenges the notion that compensatory effects between health behaviors would reduce the efficacy of multiple-behavior interventions as the number of recommendations increases. Although prior studies have noted compensatory effects between behaviors such as alcohol use and physical activity (Nigg et al., 2009) and eating behaviors and physical activity (Martin et al., 2019; Nigg et al., 2021), these studies did not examine compensatory effects in the context of multiple-behavior interventions, instead examining behavior when there was either no intervention or only a single-behavior intervention. Thus, it is possible that compensatory effects are less likely to arise in multiple-behavior interventions. As multiple-behavior interventions include specific recommendations targeting each health behavior, recipients may be less likely to use improvements in one behavior to excuse declines in another behavior.
Another finding supporting general processes of behavioral change concerns the association between changes in one behavior and changes in other behaviors. We found that the number of recommendations exerted an effect on changes in one behavior via changes in other behaviors (Table 6). As outlined in the introduction, this finding is consistent with the hypothesis of behavioral cuing, where change in one behavior cues engagement or disengagement with other behaviors. However, changes in alcohol and drug use were the strongest predictors of change in other behaviors, potentially due to reductions in substance use not only limiting cuing for other risk behaviors but also reducing disinhibition and increasing cognitive capacity to follow health recommendations.

Finally, we examined the mediator of improvement in psychological well-being to explain the effect of the number of behavioral recommendations on behavioral and clinical change. Evidence from multiple tests (Tables 6 & 7) demonstrated that psychological well-being consistently played a positive mediating role, which captures reductions in anxiety/stress, trauma, and depression, as well as improvements in coping, mood, and psychosocial quality of life. This finding suggests that simultaneously recommending multiple behavioral changes may have the power to induce a spill-over effect whereby engaging in healthy behavior in one domain leads to improvements in psychological well-being that then facilitate healthy behavior in other domains, thus reinforcing adherence to behavioral recommendations in the long term. In contrast, of the three behavior-specific mechanisms that we tested (i.e., information, motivation, and behavioral skills), none mediated the relation between the number of behavioral recommendations and behavioral and clinical change, again suggesting that it is primarily general processes of behavioral change that underly this relation.
Limitations

Several limitations of this meta-analysis are worth mentioning. First, the current research questions are limited by the kind of information searched, reported, and publicly available. Despite the overall dataset being both sizeable and comprehensive in reporting primary behavioral and clinical outcomes, we have searched and included studies that target eight domains of health behaviors and only approximately 12.4% of the included studies measured psychological well-being. As such, we chose to test mediation using measures of psychological well-being and behavioral or clinical outcomes taken at the same posttest, because using only instances in which psychological well-being was measured at an earlier posttest than the behavioral or clinical outcome would have drastically reduced the number of available effect sizes from 136 to only 45. Nonetheless, testing mediation in this fashion leaves open the possibility of a reverse causal relation (i.e., changes in behavioral or clinical outcomes causing improved psychological well-being). Furthermore, some of the included analyses are near the lower bounds of the recommended number of observations for appropriate statistical power (Tanner-Smith et al., 2016) or concessions were made in the inclusion of confounds to preserve the number of observations. Yet, given the comprehensive nature of this meta-analysis, our analyses are likely the best initial test of mechanisms in these domains of health behavior, but future costly interventions are required to verify the present findings.

Second, it is possible that the search terms we used to conduct the behavioral domain searches lacked certain keywords used in a subset of articles. Specifically, the search terms we used to identify smoking interventions did not include the word “nicotine,” while the search terms we used to identify alcohol use and substance use interventions did not include the names of common substances such as “cocaine,” “heroin,” and “marijuana.” To address this issue, we
re-ran our diet, exercise, and smoking literature search in several databases both with and without the search term “nicotine” to test whether adding this word to our original list of search terms would significantly increase the number of resulting studies. We found that, compared to using our original search terms, adding “nicotine” increased the number of search results from 1,407 to 1,414 (0.5% increase) in PsycInfo; 811 to 814 (0.4% increase) in Scopus; and 17,641 to 17,689 (0.3% increase) in Web of Science. Similarly, we re-ran our HIV, alcohol use, and drug use literature search in several databases both with and without adding the words “cocaine,” “heroin,” and “marijuana” to our original search terms. We found that, compared to using our original search terms, adding “cocaine,” “heroin,” and “marijuana” increased the number of search results from 5,102 to 5,490 (7.6% increase) in PsycInfo; 1,871 to 1,895 (1.3% increase) in Scopus; and 39,857 to 41,509 (4.1% increase) in Web of Science. Thus, although it is possible that a few relevant articles were missed during our original search, based on the above evidence, our original search terms captured the vast majority of eligible studies.

Third, the analyses targeted to investigate mechanisms of behavioral change would be optimal if well-being were fully reported in all studies. Recommended methods for tests of mediation in meta-analysis are either to calculate a joint indirect effect using meta-analytic structural equation modeling or to calculate a cross-lagged effect size that accounts for the covariance of behaviors (Cheung, 2015). Unfortunately, both approaches require complete information to calculate a variance-covariance matrix (e.g., the covariance between two different outcomes reported in each study; Cheung, 2015). Across the wide variety of studies included, the rate of correlation-matrix or covariance reporting was so low that neither of these approaches was feasible. The regression-based and path analysis models (Table 8) presented here provide less precise estimates of indirect effects over time. A better but yet unavailable approach would
be a model that helps directly test the effect of the number of recommendations on outcomes at the delayed posttest through changes in the mediators at the immediate posttest. We thus encourage researchers to report correlation matrices for all outcomes whenever possible and make de-identified data from their studies available in a more routine fashion.

Fourth, an inherent problem of all meta-analyses is that these analyses reflect effects aggregated to the sample level. Behavioral change is inherently a process that occurs within an individual, and therefore, behavioral change may not be well represented in aggregated data (e.g., Simpson’s Paradox; Simpson, 1951). The contribution of this meta-analysis, however, is to assess whether a common mechanism occurs with sufficient frequency in each intervention population to produce the expected average effects. Regardless, these results should be cautiously replicated with measures of behavioral and clinical change among individual participants.

Fifth, while our analyses suggest that the linear effect of the number of recommendations is robust to whether interventions were conducted in Western or non-Western countries and that studies in non-Western nations were associated with larger effect sizes, it should be noted that our review was comprised primarily of samples taken in Western contexts, with under 20% of the 364 included studies having been conducted in non-Western countries (see Supplement G). Furthermore, our use of English keywords and reliance on databases that cater primarily to English-language content likely precluded us from finding additional relevant non-Western studies. Thus, our findings regarding Western vs. non-Western countries should be viewed as preliminary pending replication with additional non-Western studies, and we recommend that researchers conducting future meta-analyses of this literature expand their searches to include
non-English-language keywords in Google Scholar to capture a more comprehensive and diverse set of reports in this domain.

Finally, the primary analysis in our work relied on within-group growth effect sizes. Although this approach has several benefits for answering our research questions, the lack of control through group comparison within studies is a limitation. We therefore developed the propensity score model to help make the groups receiving different numbers of behavioral recommendations more comparable at baseline and included the propensity scores as a covariate in the CHE+ models. Future meta-analytic research can explore means of utilizing between-group comparisons within studies to better understand the mechanisms associated with the impact of the number of behavioral recommendations. The network meta-analysis approach represents such a possibility (Cipriani et al., 2013).

**Future Directions**

One question in this meta-analysis is whether there is evidence of decreases in recommendation benefits as they approach a high number, as predicted if behavior-specific processes are at work and the number of recommendations is sufficiently taxing. Although little evidence of a curvilinear association between the number of recommendations and behavioral and clinical change using typical criteria has been identified, a substantial range restriction in the dataset exists, with the majority of studies presenting only zero to four recommendations and most studies topping out at five recommendations. This range of recommendations may not overtax mental resources (Cowan, 2010), or changes in several behaviors may operate efficiently as proposed. A much-needed line of research is multiple experimental tests of whether providing five or more recommendations would still produce consistent linear benefits.
Another important research direction is to elucidate differences between sequential and concurrent behavioral change and how each is associated with the mechanisms proposed here. A meta-analysis is unable to distinguish these two approaches to behavioral change given the overall large scale of time between measurements in comparison to the underlying behavioral process (i.e., 8-12 weeks), and sequential vs. concurrent change is a relatively understudied area of multiple-behavior-change research that has produced mixed results. Individual interventions have found that sequential change may be beneficial, such as a weight loss intervention that supported a sequential approach to diet and smoking cessation (Spring et al., 2004). In contrast, another that directly compared the two approaches found a small advantage for concurrent over sequential change in achieving long-term behavioral change (20.3% vs. 16.9%; Hyman et al., 2007). A more recent systematic review found only six multiple-behavior interventions that directly compared sequential vs. concurrent behavioral change, and there was no evidence strongly supporting one approach over the other when synthesizing the available evidence (James et al., 2016). The question of which approach is better and how variations in this approach relate to the number and content of behavioral recommendations is still open with multiple implications. An additional question is whether both approaches operate through the same or different psychosocial processes. Both questions invite future work, such as experimentally inducing simultaneous/sequential change and naturalistically examining each approach to behavioral change.

The current project focused on investigating two general mechanisms (i.e., behavioral cuing vs. mediation via psychological well-being) underpinning behavioral change in research on multiple behavioral recommendations. One addition to the framework worth exploring is the role of personality change in response to interventions (Roberts et al., 2017). A person acquiring and
applying behavioral skills to multiple behaviors may exhibit general changes in self-control, and examining whether or not this occurs requires further analyses to see if changes generalize to behaviors that were not addressed in the intervention. The current measure of psychological well-being encompasses items like anxiety and negative affect that are likely associated with the personality measure of neuroticism (Cattell & Scheier, 1961). Furthermore, people may respond to multiple recommendations differently depending on their personality, with more conscientious people organizing and pursuing multiple behaviors more effectively. As such, future experimental work examining individual differences in multiple-behavior interventions would be an interesting and needed line of research.

**Conclusions**

To conclude, providing multiple behavioral recommendations is positively associated with changes in behavioral and clinical outcomes. Leveraging the CHE+ meta-regression model and a comprehensive dataset, the present meta-analysis provided strong evidence that more recommendations are linearly associated with more behavioral change when the number of recommendations is within the range of 0 to 4 behavioral recommendations, which are most frequently investigated. It further identified evidence of general mechanisms, including behavioral cuing among outcome behaviors and changes in psychological well-being, to explain the advantage of multiple behavioral recommendations. Hopefully, this research will inspire further endeavors to help people achieve healthier lives and inspire practitioners to recommend multiple behaviors in critical areas of human health.
References

References marked with an asterisk have studies included in the present meta-analysis.


https://doi.org/10.1192/s0007125000292581


https://doi.org/10.1177/0018726714559752


https://doi.org/10.1001/archpsyc.1961.01710120031004


https://doi.org/10.1136/rmdopen-2019-000953


Results of a parallel randomised controlled trial. PLoS ONE, 7(12), e50032.
doi:10.1371/journal.pone.0050032.


trial of a group-based mobile intervention at nightclubs. Journal of studies on alcohol and drugs, 80(4), 423-430.


https://doi.org/10.1080/10410236.2010.521906


https://doi.org/10.1080/00221309.2014.1003029


* Christian, J., Bessesen, D., Byers, T., Christian, K., Goldstein, M. and Bock, B. Clinic-Based Support to Help Overweight Patients With Type 2 Diabetes Increase Physical Activity and Lose Weight. Archives of Internal Medicine, 168, 141-146.


https://doi.org/10.1177/0963721409359277


intervention in the primary care setting of patients with high plasma fibrinogen in the primary prevention of cardiovascular disease. BMC research notes, 5, 1-7.


Women: A Randomized Controlled Trial Effect of Diet and Exercise on the Dietary Inflammatory Index. Cancer Prevention Research, 14(1), 85-94.


Firth, J., Solmi, M., Wootton, R. E., Vancampfort, D., Schuch, F. B., Hoare, E., … & Stubbs, B. (2020). A meta-review of “lifestyle psychiatry”: The role of exercise, smoking, diet and


European multifactorial preventive trial of coronary heart disease: four-year experience.
Preventive Medicine, 12, 175-180.


randomised trials. Preventive Medicine, 89, 211–223.

https://doi.org/10.1016/j.ypmed.2016.06.012


https://doi.org/10.1093/oxfordjournals.aje.a010011


https://doi.org/10.1093/eurpub/ckt172


the chest pain observation unit: pilot randomized controlled trial. Academic Emergency Medicine, 24(8), 968-982.


https://doi.org/10.1371/journal.pone.0117015


https://doi.org/10.1001/jama.1982.03330120023025

https://doi.org/10.1080/0887044042000196737


https://doi.org/10.1001/jama.2018.0158


https://doi.org/10.1006/ceps.1999.1017


https://doi.org/10.1177/1559827610391883

https://doi.org/10.1016/j.ypmed.2008.02.001

https://doi.org/10.1080/17437199.2019.1707104

* Purcell, D. W., Latka, M. H., Metsch, L. R., Latkin, C. A., Gómez, C. A., Mizuno, Y., ... Borkowf, C. B. (2007). Results from a randomized controlled trial of a peer-mentoring intervention to reduce HIV transmission and increase access to care and adherence to HIV medications among HIV-seropositive injection drug users. JAIDS Journal of Acquired Immune Deficiency Syndromes, 46, S35-S47. doi:http://dx.doi.org/10.1097/QAI.0b013e31815767c4


* Santa-Maria, C. A., Coughlin, J. W., Sharma, D., Armanios, M., Blackford, A. L., Schreyer, C., ... & Stearns, V. (2020). The Effects of a Remote-based Weight Loss Program on Adipocytokines,
Metabolic Markers, and Telomere Length in Breast Cancer Survivors: the POWER-Remote Trial


  doi:10.1037/ccp0000096


https://doi.org/10.1037/1082-989X.7.4.422


* Stier-Jarmer, M., Frisch, D., Neuy, S., & Schuh, A. (2021). A 3-Week Naturopathic Intervention Improves HbA1c, Weight, and Quality of Life Among Overweight and Obese Adults With Type 2 Diabetes: 6-Month Results From a Randomized Trial. Alternative Therapies in Health & Medicine, 27.


vulnerable adolescents in South Africa. AIDS Care, 28(sup1), 8–15.
http://doi.org/10.1080/09540121.2016.1146213


https://doi.org/10.1097/00005650-199206000-00002


* Wilhelmsen, L., Berglund, G., Elmfeldt, D., Tibblin, G., Wedel, H., Pennert, K., Vedin, A.,
  Wilhelmsson, C., & Werkö, L. (1986). The multifactor primary prevention trial in Göteborg,

  early antenatal health promotion workshop: a randomized controlled trial. BMC pregnancy and
  childbirth, 12(1), 1-12.

  Williams, C. M. (2018). Effectiveness of a healthy lifestyle intervention for chronic low back
  pain. PAIN, 159(6), 1137-1146. doi:10.1097/j.pain.0000000000001198

  Reducing the Sexual Risk of Men Released From Jails. Journal of Correctional Health Care,
  24(1), 71-83.

  behavioral weight loss program for overweight adolescent African-American girls. Eating and
  Weight Disorders, 10, 193-203.

  among rural youth following a school-based randomized controlled trial. Preventive Medicine,
  54, 150-156.

  When it comes to lifestyle recommendations, more is sometimes less: A meta-analysis of
  theoretical assumptions underlying the effectiveness of interventions promoting multiple
  https://doi.org/http://dx.doi.org/10.1037/a0038295


https://worldpopulationreview.com/country-rankings/western-countries


https://doi.org/10.1016/j.jsat.2011.12.010


https://doi.org/10.1186/1471-2458-6-300


brief advice; a randomized trial for treatment of hazardous substance use in pregnancy and delivery. Gen Hosp Psychiatry, 34, 439-449.


http://doi.org/10.2105/AJPH.2007.126834
Table 1. Description of the Data

<table>
<thead>
<tr>
<th>Descriptive Variables</th>
<th>Variable Levels</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Report Characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Academic Affiliation ($\kappa = .93$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University/college</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td>Research center</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Hospital/health service</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>CDC</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Medical school</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Institutional Area ($\kappa = 1$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychology</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Community/public health</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Medicine</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>Intervention Characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recruitment Context ($\kappa = 1$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospital/health clinic</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td>Educational</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Non-healthcare</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>Multiple</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Delivery Context ($\kappa = 1$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health clinic</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>Community</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Schools</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Street</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Mass media</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Multiple contexts</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Communication Context ($\kappa = .97$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Face-to-face</td>
<td>87</td>
<td></td>
</tr>
<tr>
<td>Software</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Internet</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Multiple contexts</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Delivery Format ($\kappa = 1$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Groups</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>Individuals</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>Both</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Facilitator ($\kappa = .91$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Descriptive Variables</td>
<td>Variable Levels</td>
<td>Percentage (%)</td>
</tr>
<tr>
<td>-----------------------</td>
<td>------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Doctor/nurse</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>Public health educator</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>Clinical psychologist</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Community leader</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Multiple</td>
<td>18</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* \( \kappa \) = intercoder reliability for categorical variables. Academic Affiliation = the academic affiliation of the first author; Institutional Area = the subject area of the first author; Recruitment Context = the context for recruiting participants; Exposure Setting = the setting of intervention delivery; Delivery Medium = the medium used for delivering intervention; Delivery Format = the format of delivering intervention; Facilitator = the person delivering the intervention; Informational Strategies = the intervention provided informational education to facilitate behavioral change; Motivational Strategies = the intervention included motivational strategies to facilitate behavioral change; Instruction in Behavioral Skills = the intervention included teaching behavioral skills to facilitate behavioral change. The percentage values for levels of each variable were based on all the available observations from a total of 667 independent study samples.
Table 2. Covariates Included in the Propensity Score Model

<table>
<thead>
<tr>
<th>Covariate #</th>
<th>Covariate</th>
<th>Category #</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariate 1</td>
<td>Whether the intervention targeted a particular gender group</td>
<td>2</td>
</tr>
<tr>
<td>Covariate 2</td>
<td>Whether participants self-selected into the study</td>
<td>2</td>
</tr>
<tr>
<td>Covariate 3</td>
<td>Whether the intervention used attitudinal arguments</td>
<td>2</td>
</tr>
<tr>
<td>Covariate 4</td>
<td>Whether the intervention used informational arguments</td>
<td>2</td>
</tr>
<tr>
<td>Covariate 5</td>
<td>Whether the intervention taught behavioral skills</td>
<td>2</td>
</tr>
<tr>
<td>Covariate 6</td>
<td>Whether the intervention showed what to do when facing behavioral barriers</td>
<td>2</td>
</tr>
<tr>
<td>Covariate 7</td>
<td>Whether the intervention used self-monitoring prompts</td>
<td>2</td>
</tr>
<tr>
<td>Covariate 8</td>
<td>Whether the study included at least one follow-up before 6 months</td>
<td>2</td>
</tr>
<tr>
<td>Covariate 9</td>
<td>Whether the intervention aimed to improve lifestyle behaviors</td>
<td>2</td>
</tr>
<tr>
<td>Covariate 10</td>
<td>Whether the intervention aimed to improve HIV behaviors</td>
<td>2</td>
</tr>
<tr>
<td>Covariate 11</td>
<td>Whether the intervention aimed to improve substance-use behaviors</td>
<td>2</td>
</tr>
<tr>
<td>Covariate 12</td>
<td>Whether the intervention occurred in North America</td>
<td>2</td>
</tr>
<tr>
<td>Covariate 13</td>
<td>Whether the study included more than one posttest</td>
<td>2</td>
</tr>
<tr>
<td>Covariate 14</td>
<td>Time of the publication (categorized into pre-2003, 2003-2010, post-2010)</td>
<td>3</td>
</tr>
<tr>
<td>Covariate 15</td>
<td>Percentage of males in the sample (categorized into 4 percentage quarters)</td>
<td>4</td>
</tr>
</tbody>
</table>

*Note. Covariate # corresponds to those shown in Supplement H.*
Table 3. Effect of Number of Recommendations on Behavioral and Clinical Outcomes

<table>
<thead>
<tr>
<th></th>
<th>Behavioral Outcomes</th>
<th>Clinical Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Linear (n=233; k=402; s=1,501)</td>
<td>Curvilinear (n=233; k=402; s=1,501)</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.28 (0.1)</td>
<td>0.37 (0.1)</td>
</tr>
<tr>
<td>Effect-Size Variance</td>
<td>-0.42 (0.29)</td>
<td>-0.4 (0.29)</td>
</tr>
<tr>
<td>Propensity Score</td>
<td>-0.04 (0.02)</td>
<td>-0.03 (0.02)</td>
</tr>
<tr>
<td>Number of Recommendations</td>
<td>0.02 (0.01)</td>
<td>0.03 (0.01)</td>
</tr>
<tr>
<td>Squared Number of Recommendations</td>
<td>-0.005 (0.003)</td>
<td>.16</td>
</tr>
</tbody>
</table>

Random Effects

| σ₁² (studies) | 0.16 | 0.16 | 0.08 | 0.07 |
| σ₂² (samples) | 0.00 | 0.00 | 0.00 | 0.00 |
| σ₃² (measures) | 0.21 | 0.21 | 0.14 | 0.14 |

Note. n is the number of studies; k is the number of unique groups; s is the total number of effect sizes. σ₁² is the random component (variance) at the study level; σ₂² is the random component (variance) at the sample level; σ₃² is the random component (variance) at the measure level. Passive controls in the trials correspond to the number of recommendation being 0’s and active controls 1’s. *p < .05, **p < .01, ***p < .001.
Table 4. Effects of the Number of Recommendations

<table>
<thead>
<tr>
<th></th>
<th>Linear Model</th>
<th></th>
<th>Curvilinear Model</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>((n = 291, k = 501, s = 2743))</td>
<td></td>
<td>((n = 291, k = 501, s = 2743))</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(B (SE))</td>
<td>(p)</td>
<td>(B (SE))</td>
<td>(p)</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.08 (0.09)</td>
<td>.34</td>
<td>0.16 (0.09)</td>
<td>.08</td>
</tr>
<tr>
<td>Effect-Size Variance</td>
<td>0.10 (0.22)</td>
<td>.65</td>
<td>0.1 (0.22)</td>
<td>.65</td>
</tr>
<tr>
<td>Propensity Score</td>
<td>-0.02 (0.02)</td>
<td>.22</td>
<td>-0.02 (0.02)</td>
<td>.30</td>
</tr>
<tr>
<td>Number of Recommendations</td>
<td>0.02 (0.01)</td>
<td>.009**</td>
<td>0.03 (0.01)</td>
<td>.001***</td>
</tr>
<tr>
<td>Squared Number of Recommendations</td>
<td></td>
<td></td>
<td>-2e-3 (1e-3)</td>
<td>.12</td>
</tr>
<tr>
<td>Motivation(^a)</td>
<td>0.11 (0.08)</td>
<td>.14</td>
<td>0.10 (0.08)</td>
<td>.22</td>
</tr>
<tr>
<td>Clinic Context(^a)</td>
<td>0.01 (0.04)</td>
<td>.82</td>
<td>2e-3 (0.04)</td>
<td>.95</td>
</tr>
<tr>
<td>Clinical Outcome(^a)</td>
<td>-0.07 (0.04)</td>
<td>.13</td>
<td>-0.07 (0.04)</td>
<td>.13</td>
</tr>
<tr>
<td>Proportion Outcome(^a)</td>
<td>0.36 (0.1)</td>
<td>.001***</td>
<td>0.36 (0.1)</td>
<td>.001**</td>
</tr>
<tr>
<td>Duration(^b)</td>
<td>0.03 (0.02)</td>
<td>.30</td>
<td>0.02 (0.02)</td>
<td>.39</td>
</tr>
<tr>
<td>Time(^b)</td>
<td>-0.03 (0.02)</td>
<td>.08</td>
<td>-0.03 (0.02)</td>
<td>.09</td>
</tr>
</tbody>
</table>

Random Effects

\(\sigma_1^2\) (studies) \quad 0.10 \quad 0.09
\(\sigma_2^2\) (samples) \quad 0.00 \quad 0.00
\(\sigma_3^2\) (measures) \quad 0.21 \quad 0.21

Note. Motivation = intervention contained motivational strategies; Clinical Context = whether the intervention took place in a clinical or non-clinical setting; Clinical Outcomes = whether the effect size is clinical or behavioral; Proportion Outcome = The calculation of effect size for the outcome is based on proportions (vs. mean and standard deviations); Duration = the duration of the intervention program; Time = the time interval between baseline and posttest measurement. \(n\) is the number of studies; \(k\) is the number of unique groups; \(s\) is the total number of effect sizes. \(\sigma_1^2\) is the random component (variance) at the study level; \(\sigma_2^2\) is the random component (variance) at the sample level; \(\sigma_3^2\) is the random component (variance) at the measure level.
(variance) at the measure level. Passive controls in the trials correspond to the number of recommendation being 0’s and active controls 1’s. \(^a\), indicator variable; \(^b\), standardized variable. \(p < .05\), \(p < .01\), \(p < .001\).
Table 5. Exploratory Models for Identifying Sources of Heterogeneity

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Model with number of recommendations as a predictor</th>
<th>Model with all predictors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( B(\text{SE}) ) &amp; ( p ) &amp; ( B(\text{SE}) ) &amp; ( p ) &amp; ( B(\text{SE}) ) &amp; ( p )</td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.28(0.03) &amp; .001***</td>
<td>0.16(0.04) &amp; .001***</td>
</tr>
<tr>
<td>Number of Recommendations</td>
<td>0.04(0.01) &amp; .001***</td>
<td>0.03(0.01) &amp; .001**</td>
</tr>
<tr>
<td>Duration(^b)</td>
<td>0.05(0.02) &amp; .01*</td>
<td></td>
</tr>
<tr>
<td>Western Countries(^a)</td>
<td>-0.19(0.09) &amp; .05*</td>
<td></td>
</tr>
<tr>
<td>HIV(^a)</td>
<td>-0.12(0.12) &amp; .30</td>
<td></td>
</tr>
<tr>
<td>Substance Use(^a)</td>
<td>-0.17(0.08) &amp; .06</td>
<td></td>
</tr>
<tr>
<td>Mean Age(^b)</td>
<td>0.04(0.03) &amp; .13</td>
<td></td>
</tr>
<tr>
<td>Percent Males(^b)</td>
<td>0.02(0.03) &amp; .48</td>
<td></td>
</tr>
<tr>
<td>Year of Study Onset(^b)</td>
<td>0.04(0.03) &amp; .17</td>
<td></td>
</tr>
<tr>
<td>Sample Size(^b)</td>
<td>-0.14(0.07) &amp; .12</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>( \sigma^2 ) (studies)</th>
<th>( \sigma^2 ) (samples)</th>
<th>( \sigma^2 ) (measures)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0788</td>
<td>0.0731</td>
<td>0.0557</td>
</tr>
<tr>
<td></td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>0.1924</td>
<td>0.1912</td>
<td>0.1918</td>
</tr>
</tbody>
</table>

Reduction of Heterogeneity --- 7.23% 29.31%

Note. Western Country = whether the residence of participants is in western vs. non-western countries; HIV = whether the study is from the HIV domain or not; Mean Age = the mean age of the study sample; Percent Males = the percentage of males in the study sample; Year of Study Onset = the year when the study began; Sample Size = the study sample size; Duration = the duration of the intervention program. \( \sigma^2 \) is the random component (variance) at the study level; \( \sigma^2 \) is the random component (variance) at the sample level; \( \sigma^2 \) is the random component (variance) at the measure level. Passive controls in the trials correspond to the number of recommendation being 0’s and active controls 1’s. \(^a\) indicator variable; \(^b\) standardized variable. \(^*p < .05\), \(^**p < .01\), \(^***p < .001\).
Table 6. Tests of Behavioral Change Cuing Other Behavioral Changes

<table>
<thead>
<tr>
<th></th>
<th>Exercise (n = 29, k = 55)</th>
<th>Diet (n = 27, k = 54)</th>
<th>Smoking (n = 11, k = 19)</th>
<th>HIV (n = 16, k = 35)</th>
<th>AU (n = 18, k = 38)</th>
<th>DU (n = 19, k = 41)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.16 (0.15)</td>
<td>0.11 (0.08)</td>
<td>0.43 (0.54)</td>
<td>0.48 (0.26)</td>
<td>0.16 (0.08)</td>
<td>0.29 (0.14)</td>
</tr>
<tr>
<td>Effect-Size Variance</td>
<td>0.01 (0.44)</td>
<td>0.06 (0.33)</td>
<td>-1.75 (2.54)</td>
<td>-0.06 (0.85)</td>
<td>-0.35 (0.3)</td>
<td>-0.33 (0.32)</td>
</tr>
<tr>
<td>ES of Specific Behavior</td>
<td>0.27 (0.22)</td>
<td>0.14 (0.06)</td>
<td>0.02 (0.34)</td>
<td>0.7 (0.18)</td>
<td>0.48 (.01)**</td>
<td>0.35 (.02*)</td>
</tr>
<tr>
<td>Cueing Behavior Time</td>
<td>0.00 (0.09)</td>
<td>0.03 (0.04)</td>
<td>0.11 (0.09)</td>
<td>-0.28 (0.26)</td>
<td>-0.1 (0.12)</td>
<td>0.2 (0.09)</td>
</tr>
<tr>
<td>Cued Behavior Time</td>
<td>-0.04 (0.1)</td>
<td>-0.14 (0.07)</td>
<td>-0.18 (0.22)</td>
<td>0.73 (0.42)</td>
<td>0.02 (0.08)</td>
<td>-0.34 (0.07)</td>
</tr>
</tbody>
</table>

Random Effects

<table>
<thead>
<tr>
<th></th>
<th>0.20</th>
<th>0.05</th>
<th>0.29</th>
<th>0.09</th>
<th>0.04</th>
<th>0.11</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_1^2$ (studies)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_2^2$ (samples)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. HIV = human immunodeficiency virus; AU = Alcohol Use; DU = Drug Use; Cueing Behavior Time = time interval between baseline and the cueing behavior measured at immediate posttest; Cued Behavior Time = time interval between baseline and the cued behavior measured at delayed posttest; Domains include lifestyle (Exercise, Diet, Smoking) as well as one for HIV, AU, and DU. n is the number of studies; k is the number of unique groups; s is the total number of effect sizes. $\sigma_1^2$ is the random component (variance) at the study level; $\sigma_2^2$ is the random component (variance) at the sample level. Passive controls in the trials correspond to the number of recommendation being 0’s and active controls 1’s. *p < .05, **p < .01, ***p < .001. $^b$ standardized variable.
Table 7. Mediation Analysis for Psychological Well-Being, Information, Motivation, and Behavioral Skill

<table>
<thead>
<tr>
<th></th>
<th>Psychological Well-Being (n = 43, k = 91, s = 136)</th>
<th>Information (n = 16, k = 32, s = 41)</th>
<th>Motivation (n = 19, k = 39, s = 56)</th>
<th>Behavioral Skill (n = 27, k = 55, s = 74)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B (SE)</td>
<td>p</td>
<td>B (SE)</td>
<td>p</td>
</tr>
<tr>
<td><strong>Step 1: X → Y</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fixed Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.03 (0.08)</td>
<td>.73</td>
<td>0.01 (0.2)</td>
<td>.96</td>
</tr>
<tr>
<td>Effect-Size Variance</td>
<td>0.33 (0.25)</td>
<td>.20</td>
<td>0.73 (0.76)</td>
<td>.38</td>
</tr>
<tr>
<td>Time(^b)</td>
<td>0.02 (0.06)</td>
<td>.74</td>
<td>-0.11 (0.06)</td>
<td>.27</td>
</tr>
<tr>
<td>Number of Recommendations</td>
<td>0.06 (0.01)</td>
<td>.001**</td>
<td>0.05 (0.03)</td>
<td>.22</td>
</tr>
<tr>
<td><strong>Random Effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\sigma_1^2) (studies)</td>
<td>0.04</td>
<td>0.10</td>
<td>0.07</td>
<td>0.06</td>
</tr>
<tr>
<td>(\sigma_2^2) (samples)</td>
<td>0.01</td>
<td>0.06</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>(\sigma_3^2) (measures)</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>Step 2: X → M</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fixed Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.14 (0.17)</td>
<td>.42</td>
<td>-0.01 (0.12)</td>
<td>.97</td>
</tr>
<tr>
<td>Effect-Size Variance</td>
<td>0.49 (0.33)</td>
<td>.15</td>
<td>1.22 (0.62)</td>
<td>.1</td>
</tr>
<tr>
<td>Time(^b)</td>
<td>0 (0)</td>
<td>.36</td>
<td>-0.01 (0.03)</td>
<td>.67</td>
</tr>
<tr>
<td>Number of Recommendations</td>
<td>0.12 (0.05)</td>
<td>.02*</td>
<td>0.08 (0.04)</td>
<td>.09</td>
</tr>
<tr>
<td><strong>Random Effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\sigma_1^2) (studies)</td>
<td>0.01</td>
<td>0.00</td>
<td>0.02</td>
<td>0.03</td>
</tr>
<tr>
<td>(\sigma_2^2) (samples)</td>
<td>0.23</td>
<td>0.15</td>
<td>0.03</td>
<td>0.01</td>
</tr>
<tr>
<td>(\sigma_3^2) (measures)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>Step 3: X + M → Y</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fixed Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.05 (0.07)</td>
<td>.44</td>
<td>0.02 (0.19)</td>
<td>.92</td>
</tr>
<tr>
<td>Psychological Well-Being (n = 43, k = 91, s = 136)</td>
<td>Information (n = 16, k = 32, s = 41)</td>
<td>Motivation (n = 19, k = 39, s = 56)</td>
<td>Behavioral Skill (n = 27, k = 55, s = 74)</td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>-------------------------------------</td>
<td>-----------------------------------</td>
<td>------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Effect-Size Variance</td>
<td>B (SE)</td>
<td>p</td>
<td>B (SE)</td>
<td>p</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>-------</td>
<td>---</td>
<td>-------</td>
<td>---</td>
</tr>
<tr>
<td>0.22 (0.23)</td>
<td>.35</td>
<td></td>
<td>0.65 (0.76)</td>
<td>.43</td>
</tr>
<tr>
<td>Time&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.01 (0.05)</td>
<td>.8</td>
<td>-0.11 (0.06)</td>
<td>.28</td>
</tr>
<tr>
<td>Number of Recommendations</td>
<td>0.02 (0.01)</td>
<td>.12</td>
<td>0.04 (0.05)</td>
<td>.41</td>
</tr>
<tr>
<td>Mediator</td>
<td>0.35 (0.18)</td>
<td>.11</td>
<td>0.06 (0.19)</td>
<td>.79</td>
</tr>
</tbody>
</table>

**Random Effects**

| | \( \sigma_1^2 \) (studies) | \( \sigma_2^2 \) (samples) | \( \sigma_3^2 \) (measures) |
|-----------------------------------------------|-------------------------|-------------------------|
| 0.01 | 0.02 | 0.01 |

**Indirect Effect**

| | Mediation Strength | |
|-----------------------------------------------|---------|
| 0.042 | .005 | .033 | .008 |

*Note. n is the number of studies; k is the number of unique groups; s is the total number of effect sizes; X is the independent variable; M is the mediator variable; Y is the outcome variable; Time = time interval in days between baseline and outcome measure. \( \sigma_1^2 \) is the random component (variance) at the study level; \( \sigma_2^2 \) is the random component (variance) at the sample level; \( \sigma_3^2 \) is the random component (variance) at the measure level. Passive controls in the trials correspond to the number of recommendation being 0’s and active controls 1’s. <sup>b</sup>standardized variable. * \( p < .05, ** \( p < .01, *** \( p < .001. \)
Table 8. Test of Indirect Effects through Psychological Well-Being and Motivation

<table>
<thead>
<tr>
<th>Test of Indirect Effect</th>
<th>Psychological Well-Being (s = 139)</th>
<th>Motivation (s = 58)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Z / B (SE) p</td>
<td>Z / B (SE) p</td>
</tr>
<tr>
<td>Sobel Test</td>
<td>3.35 .001***</td>
<td>1.10 .27</td>
</tr>
<tr>
<td>Path Analysis w/ Bootstrap CI</td>
<td>0.05 (0.01) .001***</td>
<td>0.009 (0.008) .26</td>
</tr>
<tr>
<td>Path Analysis w/ Robust SE</td>
<td>0.05 (0.01) .001***</td>
<td>0.009 (0.006) .18</td>
</tr>
</tbody>
</table>

*Note. s is the number of effect size observations. CI = confidence interval; SE = standard error; Z = Z-value. *p < .05, **p < .01, ***p < .001.
Figure 1. Flow Chart for Study Screening and Inclusion

Summary (%) for excluding 3188 full-text reports:
1) Review, theoretical, descriptive articles, qualitative studies, medical trial (e.g., surgery and drug trial), or survey research (5.9%)
2) Presence of at least two groups (38.3%)
3) Presence of two or more behavioral recommendations in the interventions (7.2%)
4) Presence of behavioral recommendations from two or more domains (42.3%)
5) Presence of at least one behavioral or clinical outcome (1.6%)
6) Was not executed or completed, did not have baseline/follow-up statistics, or reported only mediation/secondary analyses, did not include descriptive data for calculating effect sizes (4.7%)

24,680 Lifestyle Studies Screened
15,233 HIV Studies Screened
2,152 Substance Use Studies Screened

42,065 records screened
3,552 full-text reports assessed for eligibility
364 studies (803 groups, 6,819 effect sizes) included in the quantitative synthesis

38,513 records excluded
**Figure 2. Funnel Plot of Aggregated Sample-Level Effect Sizes.**

*Note.* The plot was generated using aggregated sample-level effect sizes (to avoid data non-independence issue). The dotted vertical line corresponds to the random-effects model estimate (0.32), and the funnel shape corresponds to the 95% pseudo confidence interval.
Figure 3. Propensity Score Distributions Prior to and After Trimming

Note. Propensity score distributions are shown for each level of the treatment (i.e., number of behavioral recommendations) prior to (Panel A) and after (Panel B) trimming for the area of common support (i.e., ranges on X-axis shared by all the distributions). $k$ represents the number of trial conditions in each distribution.
Figure 4. The Effect of Number of Recommendations on Clinical and Behavioral Improvements

Note. The red regression line (with the shade representing 95% CI) is based on the CHE+ model shown in the left panel of Table 4, while keeping other covariates constant at their grand means. The effect sizes are plotted with different colors to indicate the associated number of behavioral recommendations and with different sizes to indicate weight (i.e., inverse variance).