2007

Computerized, Tailored, Theory-Based Interventions for Health Behavior Change: A Comprehensive Meta-Analysis

Paul M. Krebs
University of Rhode Island

Follow this and additional works at: https://digitalcommons.uri.edu/oa_diss

Recommended Citation

Krebs, Paul M., "Computerized, Tailored, Theory-Based Interventions for Health Behavior Change: A Comprehensive Meta-Analysis" (2007). *Open Access Dissertations*. Paper 926.
https://digitalcommons.uri.edu/oa_diss/926

This Dissertation is brought to you for free and open access by DigitalCommons@URI. It has been accepted for inclusion in Open Access Dissertations by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons-group@uri.edu.

COMPUTERIZED, TAILORED, THEORY-BASED INTERVENTIONS FOR
HEALTH BEHAVIOR CHANGE: A COMPREHENSIVE META-ANALYSIS

BY

PAUL M. KREBS

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
PSYCHOLOGY

UNIVERSITY OF RHODE ISLAND

2007

ABSTRACT

Personal behavior accounts for much of the risk associated with chronic disease, thereby providing incentive for development of interventions that offer effective prevention on a large scale. Computer tailored interventions have become increasingly common for facilitating behavior change for a number of health concerns associated with chronic disease. Systematic reviews of tailoring have been completed but a sufficient number of outcomes are now available to facilitate the quantitative analysis of overall effect sizes for this type of intervention. The present study employs meta-analytic techniques to assess the mean effect for tailored interventions focusing on four health behaviors: smoking cessation, increase in physical activity, eating a healthy diet, and receiving regular mammography screening. Clinically and statistically significant overall effect sizes were found across each of the four behaviors. Retailored interventions were found to have increased efficacy over tailored interventions based on one assessment only. The addition of counselor calls to the feedback produced greater effects initially, but these were not sustained over time when compared to retailored interventions. A nonsignificant trend was found for effect sizes decreasing over time, with the most significant drops after six months post-intervention. Mean effects did not differ by recruitment strategy and differences by theory or study group could not be adequately assessed due to sample size. Gender was the only demographic predictor associated with effect size. This analysis quantifies the effect of tailored interventions, demonstrating the ability to reach large numbers of people with effective techniques that promise to reduce chronic disease burden if implemented consistently.

DEDICATION

To my family and friends –
Your support cannot be quantified

TABLE OF CONTENTS

ABSTRACT	II
DEDICATION	III
LIST OF TABLES.....	VII
LIST OF FIGURES.....	IX
LIST OF FIGURES.....	IX
CHAPTER 1: INTRODUCTION	1
Computerized Tailored Interventions	2
CHAPTER 2: META-ANALYSIS.....	17
Benefits	17
Role in Theory-Testing	18
Procedures	19
Preliminary Analysis Issues	24
Modeling Variance	29
Effect Size Interpretation	32
Limitations of Meta-Analysis.....	35
Meta-analysis in Sum.....	36
CHAPTER 3: RESEARCH JUSTIFICATION AND PREDICTIONS.....	38
Meta-analysis and Tailored Interventions.....	38
Research Hypotheses and Predictions	39
CHAPTER 4: METHODS	41

Literature Searches	41
Effect Size Calculation.....	42
Weighting of Studies.....	44
Outliers	46
Confidence Intervals	46
Modeling Variance	47
Statistically Dependent Effect Sizes	51
Choice of Comparison	52
Missing Data Procedures.....	52
CHAPTER 5: RESULTS.....	57
General Characteristics of the Studies	57
Mammography.....	73
Diet.....	80
Exercise/Physical Activity	94
Smoking	103
CHAPTER 6: DISCUSSION.....	114
Mammography.....	114
Diet.....	116
Physical Activity.....	119
Smoking	120
Summary of Predictions	122
Further questions	131

APPENDICES.....	134
Appendix A: Coding Scheme.....	134
Appendix B: Studies Excluded from Analysis.....	137
BIBLIOGRAPHY.....	140

LIST OF TABLES

Table 1: Number of Studies Located by Behavior	58
Table 2: Studies Included for Analysis	60
Table 3: General Characteristics of the Studies	69
Table 4: Reported Health Change Theory by Theoretical Variables Employed for Feedback	70
Table 5: Studies Mentioning Use of TTM and Comparison of TTM Components Used in Tailored Interventions.....	71
Table 6: Mammography Mean ES.....	76
Table 7: Minimal Intervention v. Re/Tailored on Percent Getting Mammography.....	78
Table 8: Effect of Counselor Calls Percent Getting Mammography.....	79
Table 9: Dietary Fat Mean ES.....	84
Table 10: Minimal Intervention v. Tailored and Retailored Intervention for Fat Intake	86
Table 11: Fiber, Fruit and Vegetable Results.....	91
Table 12: Minimal Intervention v. Tailored and Retailored Intervention for Fruit and Vegetable Intake.....	92
Table 13: Minimal Intervention v. Re/Tailored on Fruit, Vegetable, and Fiber Intake	93
Table 14: Comparison of Physical Activity Outcome Measures	94
Table 15: Physical Activity Mean ES.....	97
Table 16: Minimal Intervention v. Re/Tailored on Reaching Physical Activity Criteria	98
Table 17: Minimal Intervention v. Re/Tailored on Reaching Physical Activity Criteria by Timepoint Category.....	99
Table 18: Minimal Intervention v. Re/Tailored on Increasing Physical Activity	102

Table 19: Smoking Mean ES 6-Mo Abstinence.....	106
Table 20: Minimal Intervention v. Re/Tailored on 6-Month Smoking Abstinence ...	107
Table 21: Smoking Mean ES: 24-Hr, 7-Day, and 28-Day Quit	109
Table 22: Minimal Intervention v. Re/Tailored on Short Term Smoking Abstinence	110
Table 23: Re/Tailored versus Re/Tailored + Counselor Calls Smoking Cessation ...	111
Table 24: Risk Reduction and Number Need to Treat for Mammography Screening	116
Table 25: Summary of Effect Sizes Across Behaviors and Comparisons	123
Table 26: Summary of Predictions and Findings by Behavior	124

LIST OF FIGURES

Figure 1: Effect of Re/Tailored and Re/Tailored + Calls on Mammography Screening	77
Figure 2: Regression of Percent Female on Hedges g: Dietary Fat	82
Figure 3: Re/Tailored and Re/Tailored + Calls for Dietary Fat Reduction	85
Figure 4: Regression of Percent Female on Hedges's g: Fruit and Vegetable Intake...	88
Figure 5: Effect of Re/Tailored Intervention on Fr/Veg Intake Over Time.....	88
Figure 6: Regression of Percent Female of Hedges's g, Physical Activity Criteria	96
Figure 7: Mean ES by Timepoint for Physical Activity Criteria	97
Figure 8: Regression of Percent Female on Hedges g, Physical Activity Progress ...	101
Figure 9: Regression of Recruitment Rate on Mean ES 6-mo Abstinence.....	105
Figure 10: Regression of Retention Rate on Mean ES 6-mo Abstinence.....	105
Figure 11: Mean ES for Smoking Cessation Interventions by Timepoint.....	112
Figure 12: Long Term Effect for Smoking Cessation for Retailored versus Counselor Calls	113
Figure 13: Regression of Timepoint on Log Odds for Short Term Smoking Cessation	121
Figure 14: Tailored v. Retailored Mean ES over Time, Smoking Cessation Short-Term Quit.....	126

CHAPTER 1: INTRODUCTION

No doubt exists that chronic health problems negatively impact not only on personal well-being but impose a significant burden on society. In 2005 alone, the Centers for Disease Control (CDC) estimated that “Chronic diseases account for more than 75% of the nation’s \$1.4 trillion medical care costs” (Centers for Disease Control, 2005). Accounting for these costs are the chronic diseases that are responsible for nearly 70% of deaths in the US. Heart disease, cancer, and stroke comprise the leading causes of death, with these attributable primarily to smoking, poor diet, lack of exercise, alcohol use, and infection. These estimates thus suggest that the majority of chronic health problems can be prevented by decreasing rates of smoking, increasing physical activity, and improving dietary behaviors. Indeed, the Institute of Medicine suggests that over 60% of chronic disease is attributable to personal choices (Institute of Medicine, 2001). Minorities especially bear a disproportionate share of the chronic disease burden. Death from cardiovascular disease in 1998 was about 30% higher among African American adults than among white adults and the disparity continues to increase. Diabetes, an indicator of other health problems such as heart disease and obesity, is 70% higher in African Americans than whites and 100% more prevalent among Latinos than whites (Centers for Disease Control, 2005). Effecting change in chronic disease rates thus requires interventions aimed at influencing individual behavior.

Billions of dollars have been spent developing programs aimed at preventing chronic disease through health behavior change. Recognizing that the etiology of disease reaches beyond traditional medicine, the field of psychology has been lending

its knowledge and resources to studying behavior change. Over the past 30 years the field has produced numerous theories and has developed thousands of interventions for smoking cessation, physical activity promotion, dietary change, alcohol reduction, and multiple other behaviors that could contribute to disease prevention. Even with the proliferation of theory and interventions, changing behavior remains a difficult task. For example, smoking cessation interventions at their most costly incarnations produce at best a 30% cessation rate. Individual, clinic-based programs incur significant cost and necessitate additional training for busy practitioners. Interventions that could reach greater numbers of people with less practitioner training and fewer costs while remaining effective are needed to reduce overall disease burden.

Computerized Tailored Interventions

Changing the behaviors of a population to reduce risk of chronic disease requires state-of-the-science interventions. Both communicating a health message and motivating people to take action are necessary to accomplish health promotion, but the methodologies with which these are accomplished change rapidly. Print communications are a principle method of informing populations about such health behaviors. From brochures found in hospitals and medical offices to the growing number of Internet sites, to articles in popular magazines, information regarding health is prolific. This variety of modalities can be categorized into three main forms: generic, targeted, and tailored.

The first form, generic messages, such as brochures and health pamphlets, are the most abundant means of print health information. Hospitals, agencies, and foundations produce thousands of pamphlets on any variety of health topics.

Information-based Internet sites can also be considered in the category of generic interventions. Pamphlets and websites may communicate a health message but do not match their message to characteristics of the prospective consumers. They often attempt to include as much information as possible, aiming to provide something of interest to every reader. This leaves the consumer, however, to wade through the information and provides no guidance on what advice is most personally relevant.

Targeted interventions, such as a mass mailing to a population with diabetes, may increase message specificity, but cannot address variations among subgroups. Differing patterns of needs may exist along various gender, ethnic, and social lines that such general messages cannot address. As Kreuter et al. (2000) point out, traditional public health campaigns largely follow a health publicity model, operating from the belief that knowledge leads to behavior change. Such a perspective led to the use of television to reach large numbers of people with factual health information. Little evidence exists that even the most sophisticated anti-smoking commercials have an effect on cessation and prevention rates. Without a two-way feedback loop, such costly and visually appealing ads lack a degree of personal relevance. For example, an outcome study of four consecutive Dutch mass media campaigns to reduce dietary fat found no effect, possibly because this modality does not allow for assessment and individual feedback, a necessary component for change (Brug, Steenhuis, van Assema, Glanz, & DeVries, 1999). This data illustrates that complex behaviors require knowledge of cognitive and behavioral patterns on an individual level (Kreuter et al., 2000).

The third category of health communications, tailored interventions, provides for individual assessment and feedback and is thus becoming an increasingly common method of facilitating health behavior change. According to Kreuter et al. (2000) "Tailored health promotion materials are any combination of information and behavior change strategies intended to reach one specific person, based on characteristics that are unique to that person, related to the outcome of interest, and derived from an individual assessment." Given that tailored messages are composed of a combination of individual information, assessment, and feedback, infinite derivations from many theoretical perspectives are possible. Such diversity offers an opportunity for creating unique and effective interventions that is both exciting and challenging for the field. Each component of a tailored intervention from assessment to feedback must be planned and carefully considered.

Assessment

The possibility of tailoring exists when two conditions are met: when variation in the audience exists and when complex outcomes are possible. Matching audience variation with personally relevant messages requires assessment. According to Kreuter's definition, tailoring should be "based on characteristics that are unique to that person, related to the outcome of interest, and derived from an individual assessment." Psychology, more so than any other field, has adopted the study of individual characteristics. From the questionable goals of Galton's eugenics to Binet's attempts at improving children's education, to the present day study of personality by McCrae and Costa, the study of individual differences has comprised a major theme of research (Klie, 1997). Systematic study of difference has accordingly required the

creation of assessment instruments, either conducted through self-report or observation. Despite this proliferation of assessment, psychology historically has focused its lens mainly on the study of personality differences, to the neglect of systematic and theory-based assessment of mental diagnoses and other psychological realities such as health behavior.

As previously stated, psychology has a strong assessment tradition, but one that has rarely been used to inform interventions. Since assessment has been the realm of personality researchers and to a lesser extent, clinicians, psychology and related fields fall upon flawed clinical decision-making when choosing treatments. In a 1983 survey of psychologists, Norcross and Prochaska found that research findings exhibited a weak to moderate influence on practice and that outcome research ranked 10th among other factors, such as supervisory influence, in affecting a psychologist's choice of treatments. Twenty years later, Kopta et al. (1999) argue that psychology has no empirical norms for how, when, and why patients progress. Treatment planning and even manualized treatment does not specify the most important treatment variables or intervene specifically upon them because, surprisingly and regrettably, we have struggled to define them. For example, the Hawaii Integrated Healthcare project planned manualized therapies and outcome measures, but pre-treatment assessment operated from clinical impression (Laygo et al., 2003). Traditional instruments such as the MMPI may predict that a patient may have more anger than the normative population, but a therapist still has no basis for ipsative comparison, nor does the instrument provide any advice on what variables require focus. Goldfried (1980) has called for the delineation of therapeutic change principles with the hope of defining a

set of empirically based principles to guide practice. Arising from reasons including sheer difficulty, theoretical differences, and entrenchment in tradition, grounding intervention in assessment has been an elusive target.

Since the psychological tradition provides little guidance, state-of-the-science interventions require creative solutions to position psychological assessment as a foundation of behavioral health intervention. Llewelyn and Kennedy (2003), for example, describe a three-dimensional model of psychological interventions for health behavior: problem, assessment, and intervention. In this model, the ten most common health problems interact with assessment, which interacts with intervention services. This occurs in the context of individual, family, provider, and socio-cultural factors. Such a model appears simple, but breaks new and vital ground in the search for effective practices to reduce disease burden.

Assessment and Health Communication

Each level of tailoring necessitates a differing degree of assessment: (1) *Generic tailored* messages contain as much information as possible, allowing people to decide what to take from them. Such a modality requires minimal assessment, as little as asking if someone smokes or not; (2) *Personalized* communications simply use a person's name in a generic message, thus requiring little assessment; (3) *Targeted* generic communications are based on "market segmentation" for a specific population. They continue, however, to assume homogeneity in the population. Targeted interventions can entail some assessment, such as determining stage of change, from which a person could be sent a generic change manual; (4) *Tailored* communications are a "combination of strategies and information intended to reach

one specific person based on characteristics that are unique to that person, related to the outcome of interest, and derived from an individual assessment” (Kreuter et al., 2000, p. 277). Obviously, the more assessment done, the more individual the feedback will become. Learning theory has determined that feedback is essential for reinforcing and correcting behavior. Petty and Elster (1981) propose this occurs through “elaboration likelihood” such that people process information more actively if they find it personally relevant. Elaborated messages are thought to lead to more change by eliminating irrelevant information, enabling a person to attend to the most salient points, which may then result in reconsideration of behaviors and, eventually, to change. Since each level of communication requires more assessment, the main question remains how to determine the most salient variables upon which to intervene.

Theory-Based Intervention

Assessment of variables shown to produce change can guide treatment, but what forms can that intervention take? Can valid assessment of individual factors occur and can treatment be matched to each individual based on its findings? Can the factors that create change in the process of individual therapy be applied on a broader scale? These questions form the central core of applying the best techniques from individual change theory to the public health arena.

If assessment is to guide intervention, variables that effect change must be identified and assessed. Some promising veins of research have been developed to aid intervention planning. Various health behavior change theories such as the Transtheoretical Model (Prochaska & DiClemente, 1982), Health Belief Model (Rosenstock, 1966), and Theory of Planned Behavior (Ajzen, 1985) have attempted to

determine variables that underlie health-related behaviors. This vein of research assumes that, once discovered, intervening on these variables will lead to behavior change. For example, the Health Belief Model (HBM) asserts that susceptibility to illness, severity of an illness, and barriers and benefits of a suggested action influence whether a person will carry it out. A meta-analysis of the variables proposed by the Health Belief Model found that the barriers and benefits of a behavior are more predictive than susceptibility (Becker & Rosenstock, 1984). The model, however, has been criticized on two main grounds: that it focuses on rational thoughts to the exclusion of emotional factors and that it assumes people actively process health information (Ogden, 2000). The Theory of Planned Behavior (TPB) added the concept of personal value to the rational conceptualizations of the HBM. This theory proposes that several beliefs influence behavioral intentions, defined as “plans of action in pursuit of behavioral goals” (Ajzen & Madden, 1986). Intentions are formed from a person’s attitude toward a behavior, social norms, and perceived control, also known as self-efficacy. The TPB has received criticism for neglecting to propose and research causality among its variables, but nevertheless has successfully been employed to inform interventions (Norman & Conner, 1995).

The Transtheoretical Model (TTM) began with examination of naturalistic change itself in the hope of creating change through specific interventions. Research proposed and eventually supported the concept that people go through five stages of change from not thinking about change to maintaining change. These changes occur through the action of ten change processes that were derived from many models of psychology. The model has also incorporated variables common to other theories such

as benefits, barriers, and self-efficacy. The TTM has received criticism regarding whether or not change occurs in discrete stages and regarding its applicability to a variety of health behaviors (Ogden, 2000). Each theory of behavior change represents a promising intervention strategy, but as Prochaska (1999) argued, an intervention requires inclusion of the *strongest predictive variables* of not only change, but process and retention as well.

Population-Based Methodology

Theoretically, interventions that can produce change among a large number of people will have broad health impact when measured in terms of cost and general health. Interventions that reduce relative risk of a developing a disease, such as smoking cessation for reducing rates of lung cancer, can certainly help improve health outcomes, but does not specify how common a risk factor is in the general population. Population attributable risk measures the proportion of *excess* disease attributed to a risk exposure, whether unhealthy eating, lack of exercise, or smoking (Rychetnik, Frommer, Hawe, & Shiell, 2002). This statistic thus shows the potential for a prevention program to increase life expectancy, quality of life, cost, etc, if exposure to the risk factor is reduced or eliminated. Attributable risk refers to the effects of *disease* on a population, whereas the term impact can be used to refer to the effects of an *intervention* in reducing disease in a population. Impact becomes a vital consideration when public health is concerned. The idea of impact can be exemplified in the following equation: $\text{Impact} = \text{Effect Size} \times \text{Reach}$ (Prochaska & Velicer, 2004; Glasgow et al., 2006). This equation suggests that if an intervention has a large effect size and is extremely effective in helping people exercise, but can only reach two

people per year, it will have little impact on overall health. If, on the other hand, an intervention is moderately effective and can reach 10,000 people it will have great impact on decreasing overall healthcare costs and in improving health in a population. Individuals may change behaviors only slightly, but small changes are magnified when considered socially. For example, Prochaska et al. (2001) state that reducing two unhealthy behaviors reduces healthcare cost by \$2,000 per year. Cost in itself is not a value in considerations of health, but can be a predictor of increased quality of life for individuals since decreased cost could indicate less treatment seeking. The U.S. Department of Health recently reported that healthcare costs rose at their fastest rate in 15 years and will consume 20% of all spending by 2015, a large portion of which will be subsidized by the government, that is, the population as a whole (Poissal et al., 2007). Intervening on a problem with low relative risk, but high prevalence can have broad population impact and establishes the logical ground for population-based interventions.

According to Peters and Elster (2002), population based medicine involves defining a population, identifying needs, delivering services, assessing impacts, and providing feedback. Population-based intervention accounts for the fact that since health problems are situated socially, interventions that concentrate on groups are necessary to impact individuals (Jeffrey, 1989). Population intervention also allows directing limited resources, meeting preventive guidelines, and mitigating economic disparity in healthcare availability. This perspective enhances a biomedical model directed at curing specific diseases. Preventive interventions reduce the need for care, inappropriate demand for care, inappropriate use, and poor delivery of care (Peters &

Elster, 2002). Integrating population-based medicine and insights from individual treatment may create unprecedented impact on a society's health.

As an example of an application of population-based medicine, the American Medical Association (AMA) has recently realized the need to move toward population prevention and risk management. They have produced a document entitled "A Primer on Population-Based Medicine" (Peters & Elster, 2002) in which the authors delineate a possible means for clinicians and medical managers to integrate population assessment and intervention into practice. In their model, preventive service delivery would entail five steps to be employed in a clinic or HMO setting:

1. *Define a population and organize it by the 10 leading indicators of health risk.* Clinical systems should develop continual assessment of their patients' health risks, which can be sorted using the leading indicators of health risk for intervention purposes.
2. *Create information systems.* Computer systems should expand beyond billing to include systematic health risk records. Such a database could keep track of risk information upon which to recommend prevention programs. This would meet Heller and Page's (2002) goals of "developing a methodology of similar impact to that of evidence based medicine to provide an evidence base for population as well as individual health problems."
3. *Identify and prioritize into patient groups.* With a health risk computer system, conditions most prevalent in a clinic population could be identified. For example, if the clinic population showed diabetes risk, an

intervention could be developed for prevention in that population.

Statistical modeling could also identify risk factors and characteristics related to target conditions.

4. *Identify interventions.* Once people are identified at risk, they could be proactively recruited for an intervention through flagging a file for physician intervention, invited to enroll in an online program, or automatically sent an intervention packet.
5. *Adapt the system.* Taking best practices as process seriously, the procedures would be constantly evaluated and improved for identification, assessment, and intervention. Patients may want to fill out assessments while in a waiting room and have feedback to take home or they may want Internet based behavior change. Systems should take the suggestions of effectiveness research and measure variables in addition to disease outcomes.

This model serves as a possibly groundbreaking population intervention strategy for integrating prevention into primary care. It follows the recommended, but neglected, practice of using assessment as a tool to give reliable feedback for intervention. Behavior change theory and intervention techniques would fit perfectly into this model. Health promotion means giving people control over their own health. The idea now requires systems to implement it.

Population Methodology in Practice

The public health, insurance, and medical industries have attempted community-based and other interventions to promote behavior change, but with little

to no success. These education-based interventions are based upon generic or, at best, targeted methodologies and therefore produce little behavior change, even though most people believe behaviors often intervened upon, such as smoking, are harmful (Weinstein, 1984).

Public health messages have difficulty affecting an individual's decision-making since they do not convey personal urgency (Jeffrey, 1989). Successful interventions need to apply evidence-based behavior change strategies while providing elaboration and engaging the feedback loop.

The move toward empirically supported treatments in individual therapy and medicine is occurring for public health interventions as well. Evidence-based public health involves development and implementation of effective programs aimed at improving the health of a population at high risk. Heller and Page (2002) advocate reconceptualizing evidence-based medicine as "evidence for population health." Such a conceptualization means that interventions should follow requirements including ease of administration, low respondent load, reliability and validity (Velicer et al., 2000). The area of tailored communications meets the need for cost-effective, efficacious, and practical population interventions that hold participants' attention through individualization of messages. Many authors recommend tailoring as a possibility for population intervention. Glasgow et al. (2002) believe interventions need to be tailored to personal variables and organizational climate, overall making it more likely that results will replicate and generalize. Tailoring promises to meet the need for population interventions based on reliable evidence and assessment.

Reach and Implementation

A vital consideration that influences the development of a theory-based intervention is its intended reach. Innumerable interventions based on theoretical constructs are possible and can be created in variety of ways. Assessment-based, tailored interventions can form the core of a clinic-based, kiosk interaction or be part of a systematic intervention program conducted by a health insurer for a subset of thousands of members. Thus the intended audience of interventions can be large to small. Reach also interacts with format of the interventions. An office kiosk-based intervention can provide assessment and feedback entirely through a computer screen or could have a printout feedback component. A larger reach program could be conducted through telephone or mail survey methodology and participants could receive feedback via mail. Additionally, the increasing availability of Internet connectivity permits assessment and feedback from a participant's home or office.

Computerized assessment and intervention is the ideal format for delivering health interventions at low cost to a large number of people. Such communications can occur in any medium – print, Internet, and phone. If relatively simple and valid assessment instruments are developed, customized health messages can be provided on a large scale. The ability of tailoring to reach large groups, such as a population of HMO subscribers, increases the impact of interventions far beyond the possibilities of one-on-one counseling interactions. In fact, one study found no dose/response relationship, suggesting more contacts do not improve outcomes (Velicer, Prochaska, Fava, Laforce, & Rossi, 1999). In addition, Prochaska and colleagues (Prochaska et al., 2001) found that counselors did not add to computer-based tailored interventions.

If tailored communications can be made practical, they can reach many people and cost effectively produce behavior change. Brug et al. (2005) have suggested that wider distribution through non-print media such as the Internet allows wide distribution at low cost.

Population-based interventions following the AMA model have already been developed and carried out. Interventions based on the Transtheoretical Model have used employers, HMOs, or clinics to provide telephone lists for screening employees and family members for health risks. These people can be contacted by letter, email, and through use of telephone interviewers, recruited for the study, and assessed on theoretically relevant variables. The information gained can be computer-processed and mailed back to the participant in the form of communication tailored to theory-based variables. Internet-based interventions can provide immediate feedback and also link to the feedback database. Participants can be assessed in the future through mail-in forms or additional phone conversations and mailed information tailored to their change since the last assessment. This system creates the ideal therapy situation in which feedback and intervention are specifically tailored to assessment. The research behind the theory and intervention assures, to a good extent, that the variables are salient and the feedback empirically validated – a situation that rarely occurs. For example, if someone scores high on the consciousness raising process of change, she gets reinforcement, and if low, specific tips on how to improve. Other modalities for intervention and assessment include Internet-based interventions, in which participants, such as HMO members, can access assessment and feedback from home.

Automated telephone systems have been developed and are currently being tested for assessment and feedback as well.

Various forms of tailored health messages have been developed and deployed for interventions. Tailoring has been shown more effective than other forms of health communication for smoking (Prochaska et al., 2001; Prochaska, DiClemente, Velicer, & Rossi, 1993; Strecher, Kreuter, Den Boer, & Kobrin, 1994), reducing fat intake (Brug, van Assema, & de Vries, 1996; Campbell, DeVellis, Strecher, & Ammerman, 1994), increasing physical activity (Bull, Kreuter, & Scharff, 1999; Kreuter & Strecher, 1996; Marcus et al., 1998a), and getting mammograms (Skinner, Strecher, & Hospers, 1994). A variety of approaches, however, fall under the title of tailored interventions. To highlight a few distinctions, tailored interventions differ in terms of the theory upon which they are based, feedback modality, amount of assessment, variables intervened upon, type of delivery channel, and dose of intervention. This project aims to research these divergent methods of tailoring to provide greater insight into an optimal formula that would increase intervention effectiveness. Helping to determine a combination of variables (e.g. amount of tailoring, format, and theoretical constructs) that produce optimal tailoring will enhance the efficiency and impact of health behavior interventions.

CHAPTER 2: META-ANALYSIS

Given the criticisms of null hypothesis testing and the increasing amount of often-discrepant research in most fields, analysts are increasingly relying on meta-analysis to provide clearer bases for inference. Meta-analysis, a term first coined in 1976 by Glass (1976), describes a synthesis tool that pools data from many different studies asking similar questions. Not only does it bring results of different studies together, but techniques associated with it can be used to advance theory. Meta-analysis allows more precise estimates of treatment effects, helps to explain heterogeneity among studies, aids in resolving conflicting results, and can be used to establish grounds for research-based policies.

Benefits

Meta-analysis allows a precise estimate of treatment effects since it uses a continuous measure of outcome rather than the dichotomous ‘significant’ or ‘non-significant’ declarations traditionally used to report results. Low powered studies, far too common in psychology, yield false nonsignificant results at unacceptably high Type II error rates, leading to many laments about lack of progress in the field (Schmidt, 1996). Even Pearson as far back as 1904 predicted, “Many of the groups are far too small to allow of any definite opinion being formed at all, having regard to the size of the probable error involved” (Pearson, 1904, as cited in Egger, Smith, Schneider, & Minder, 1997). Given the artifact of significance and sample size, differences in statistical power create conflicting results among studies. Studies with high power and with low power may share the same effect size but show a different significance testing result. Thus, narrative reviews that simply count significant results

mislead the field from the true results of a study. This can also lead to unnecessary investigations into moderator variables to explain why certain studies were and were not significant (Schmidt, 1996). Meta-analysis can be used to resolve such conflicting results. When comparing many studies, it allows *standard objective measure* of outcome instead of narrative descriptions such as “some evidence.” The technique can also control for study-level sampling and measurement error. Under certain conditions, meta-analysis as a technique can increase the statistical power of finding significant overall effect size by reducing standard error (Cohn & Becker, 2003). Examination of confidence intervals from a series of studies often reveals if they estimate the same population parameter (Schmidt, 1996).

Meta-analysis helps to explain the heterogeneity found among studies. The differences among effect sizes of studies may follow some pattern, such as by gender or ethnic group. Using moderators in the analysis enables disentangling of method, substance, and error (Marsh, Johnson, & Carey, 2001). A researcher can use these moderators as a priori hypotheses going into the analysis. Although not often used, Shadish (1996) argues meta-analysis permits identification of mediators such as how peer pressure mediates effects of treatment on alcohol behavior. The technique can also compare the methodological quality of studies to determine if certain conditions had effects on the outcomes (Lipsey & Wilson, 2001).

Role in Theory-Testing

Meta-analysis can have positive effects on a field in general. Psychology especially, and other fields as well, are awash with conflicting evidence from non-comparable studies. This is a troubling revelation since research forms the basis for

policy and action. Practical applications desperately need consensus from the research domain to proceed. Meta-analysis can help bring resolution to uncertainty and suggest policy imperatives. Inherently, the procedure is a statistically and interpersonally less biased review method than systematic review (Egger et al., 1997), one drawback of the Cochrane review system. Bias is controlled in a manner not feasible in a systematic review if the analysis employs systematic implementation without reference to study title or authorship. Statistically, the procedure controls for artifacts of sample size, design, and error. When a meta-analysis is compiled for a particular topic, more studies can be added as they arise, leading to a cumulative meta-analysis database. This process can identify when an effect first showed up or when something changes an effect. If an analysis compiles results and finds iatrogenic or small effects, it can prevent waste on continued studies. The act of compiling studies also shows gaps and weaknesses in the literature. Meta-analysis permits a solid overview of a research field and can move a field toward the “big picture.”

Procedures

The Literature Search

In the preparatory phase, the most important consideration is that of forming a solid research question. It should be specific enough to find relevant research, but broad enough to be useful in answering the question at hand. An initial delve into the literature can help specify an accurate and realistic question. Studies should be conceptually similar to ensure the validity of conclusions. Once the question is established, the analyst begins locating and retrieving data/papers. This process too requires explicit criteria. Since research often begins in electronic databases, keywords

should be documented and modified as the researcher progresses and gains familiarity with the literature content. Distinguishing features found in abstracts can also be used. Searching for a particular demographic (women, minorities) or research designs (RCTs, quasi, etc) may help limit searches. The analyst also considers the relevance of cultural/linguistic range and time frame of the studies. Publication type can also influence inclusion criteria since many meta-analyses only employ peer reviewed journal articles. The stricter the criteria, the more credible studies will be included, but this results in smaller, N's, loss of data, limited generalizability, and inflated effect sizes (Egger et al., 1997). Once the analyst specifies criteria to narrow searching, he begins locating studies. Searching usually proceeds first in numerous electronic databases. These are often area-specific such as Medline for medicine and Psychlit for psychology. The analyst must use multiple databases since articles are listed in some but not in others. Also, searching should proceed at multiple institutions since libraries purchase different levels of database detail. In addition to primary database searches, reviews, references, journals, conferences, authors, and government agencies can provide references. When retrieving studies every effort must be made to limit bias. For example, bias can enter a study if dissertations are left out systematically. Library loans, reference librarians, government agencies, APA, and professional organizations must be utilized to find a representative sample of studies. Letters to prominent authors in a field and research organizations should be used to locate studies. Such efforts can limit the publication bias for significant results with larger effects (Begg, 1994; Lipsey & Wilson, 1993; Stern & Simes, 1997). Given these necessities the analyst needs to schedule sufficient time for data collection.

Identifying Studies

Since the researcher will want to specify initially the broadest criteria possible, the searches at first will identify a large number of studies. Being too specific with a electronic search may erroneously limit the sample of studies identified. Unexpected titles and phrasings may become apparent only when a broad sample of studies is examined by the researcher. This entails a great deal of time and concentrated work, but will decrease bias in retrieval. To aid in this effort, criteria need to be specified to determine which studies to obtain in full-text format. When examining titles and abstracts, the analyst will consider: (1) Is the study relevant to the research question? (2) Does the study include the variables of interest? (3) Does the study employ the selected methods (i.e. RCT, pre/post design, case control, etc)? (4) Does the study fall within the selected timeframe for the analysis? Studies that meet these criteria should then be downloaded or requested in full-text form for further review.

Data Extraction and Variable Coding

Studies then require coding into a database for analysis. Software options should be considered from the start as incompatibility may arise. Programs exist for meta-analysis, such as Comprehensive Meta-Analysis and Easy MA, and each have pros and cons in terms of data modeling, data entry, display, and analyses offered. Separate programs can be used such as databases for entry and statistics programs for analysis. Commonly used programs, such as Microsoft Access and Excel, can also prove to be flexible programs for entry and analysis, along with Reference Manager for the study database. Data can then be imported into standard analysis packages such

as SPSS or SAS for which meta-analysis macros have been written that result in accurate parameter estimates.

Coding should proceed according to a coding manual of variables that is also open to changes as analysis proceeds. The coding itself should be done by two independent coders who have training in the specific content of the literature, in procedures common in the content area, and in meta-analysis techniques. To control for bias, they should be blinded to names of authors and journals. Quality should be reviewed periodically and any questions documented. Inter- and intra-coder consistency should be measured. After a time, a subsample can be drawn and recoded, comparing them with percent agreement or by using inter-rater reliability statistics. Since reporting of some variables is often poor, coders can give a confidence rating on the most important variables.

Choice of variables to code is an important decision since it determines what analyses can be done later, especially in terms of moderator analyses. Overall study descriptors and effect sizes need to be coded. Descriptors include date, form of publication, authorship, population, methods and procedures, variables specific to a field, and methodological soundness.

Effect Size Entry

Effect sizes (ES) can be determined directly or estimated from information. Statistical information required includes: timepoint, sample effect size, subsample effect sizes, means, standard deviations, sample sizes, correlations, and significance levels (Rosenthal, 1995). An effect size quantifies, in either direction, the magnitude of a relationship. As such, it estimates the effect of an independent variable on a

dependent variable. Overall, the same statistic must be used across comparisons. If they cannot use the same statistic, then separate analyses must be done. Also, effect sizes must be independent so as not to affect statistical tests. Three types of effect sizes exist: mean difference, association, and multivariate.

Mean difference effect sizes are reported as Cohen's d , Hedges g or Glass's Δ . All are mean differences divided by standard deviation. Mean differences are either one-variable or two-variable relationships. One-variable relationships include mean, median, mode, or proportions. An example would be comparing scores on two measures of the same construct (Lipsey & Wilson, 2001). Two-variable relationships are either pre-post, or control group comparisons. Standardized ES's are used when dependent variables are not operationalized the same for both groups.

Beside mean differences, contrasts can appear as proportions for dichotomous outcomes. These are not preferred because effect size values depend on where the proportions fall between 0 and 1. Odds-ratio is an improved method of reporting dichotomous outcomes.

An association effect size is any two-variable relationship and is reported as Pearson's r or as Z according to Fisher's r to Z transformation. Continuous and dichotomous variables require a point biserial correlation, while two dichotomous variables require an odds ratio. Multivariate effect sizes are provided with analyses such as for multiple regression, factor analysis, and structural equation modeling. This creates problems because regressions estimate different parameters and standard errors cannot be computed. The analyst must synthesize statistics (correlations) and meta-

analyze them by doing multivariate analyses on synthesized matrices. Unfortunately few studies report full r matrices.

Preliminary Analysis Issues

Data preparation procedures proceed in ways similar to traditional analyses – by examining the distributions of data. Analysis of the mean effect, range of effect sizes, sample sizes, outliers and missing data all must be attended to before analysis ensues. Histogram and stem and leaf displays are excellent for showing central tendency, variability, and normality and diagnosing skewness and outliers. Adjustments to effect sizes often must be made at the level of the study. The researcher must weigh the pros and cons of these procedures because corrections of some biases can increase others. Analyses can be reported with and without adjustments and compared. For example, measurement error correction increases sampling error. Most often the analysis uses correction for attenuation due to unreliability, which occurs when sample effect sizes have a smaller range than the population. Additionally, biases specific to meta-analysis must be dealt with. These are publication bias, sample size bias, and artifact, or measurement, biases.

Sample Size Bias

One main strength of meta-analysis is the ability to achieve higher power to detect population differences from studies with small sample sizes. Inevitably in a meta-analysis, studies of various sample sizes will be included. Simply taking the mean effect size from these studies does not account for the differing error variances associated with sample size. Larger samples theoretically result in more accurate estimates of the population mean and thus should receive greater weight in the pooling

of effect sizes. Before pooling estimates, each effect size is weighted by multiplying the effect by the inverse of its variance, which helps correct for error variance associated with sample size. Calculating a mean effect size always involves weighting individual effect sizes by their reliability.

Missing Data

Missing data can bias meta-analyses as in any other study. If data are missing nonrandomly, it usually suggests systematic bias. In the case of meta-analysis, bias enters when studies with small effect sizes are included less often than studies with larger effect sizes. If a study reports results as nonsignificant without providing statistical specifics, the effect can be included as zero. This is a conservative procedure, however, and may nullify the aim of including underpowered studies.

Measurement Bias

Effect sizes are often dependent on outcomes measured by various testing instruments. The greater the unreliability of the measurement instrument, the more the effect size will be underestimated. Mean difference effect sizes are weighted by the squared inverse of their standard error (SD of the sampling distribution), their “inverse variance weight.” Odds ratios are corrected by taking the natural log and correlations with Fisher’s Z. Hunter and Schmidt (1990) described other adjustments to approximate ideal conditions such as unreliability, restricted range, dichotomized continuous variables, etc. Their correction procedure for measurement error is expressed in equation 1:

$$\delta_{ES'} = \frac{\delta_{ES}}{\sqrt{r_{yy}}} \quad (1.)$$

where ES = the observed (attenuated) effect size estimate, ES' = the disattenuated effect size estimate, and, $r_{yy'}$ = the reliability of the dependent variable measure, which is estimated using the reported value of the scale's internal consistency coefficient alpha. This then depends on whether the reliability of the instrument is reported, which often is not done in outcome papers, thus requiring the researcher to obtain instrument development studies.

Publication Bias

Related to the problem of missing data is that of publication bias. The theory of meta-analysis assumes that a representative or even comprehensive sample of studies has been included that show both significant and nonsignificant results. It has been shown that studies with nonsignificant findings are often not published, whereas a tendency exists to publish results of small sample size studies that result in large effects (Lipsey & Wilson, 1993). This bias of publication has been termed the 'file drawer' problem, referring to the fact that results of many studies remain unknown due to difficulty of publishing nonsignificant findings.

Various methods have been developed to assess for publication bias. The first method of studying publication bias is to plot effect sizes by their standard errors, forming what is known as a funnel plot. Studies with smaller error variances will cluster near the top of the plot, and studies with larger variance will fall out near the bottom, dispersing to the right and left of the mean. If more studies fall near the bottom and to the right of the mean, one can assume that a bias exists for publishing these small N studies with large effects. For instance, funnel plots can be difficult to interpret and asymmetry found in the funnel plots may be due to the presence of

heterogeneity of the studies rather than publication bias (Egger et al., 1997; Sterne, Gavaghan, & Egger, 2000) or may be due to both heterogeneity and publication bias (Pham et al., 2001). This is merely a visual analysis tool and others employ statistical techniques. Duval and Tweedie (2000) devised a technique for imputing values assumed to be missing in the funnel plot and allows calculation of a mean effect adjusted for publication bias.

Egger's linear regression method quantifies the bias captured by the funnel plot. In the Egger test (Egger et al., 1997), the standardized effect (effect size divided by standard error) is regressed on precision (inverse of standard error). Small studies generally have a precision close to zero, due to their high standard error. In the absence of bias one would expect to see such studies associated with small standardized effects and large studies associated with large standardized effects. This would create a regression line whose intercept approached the origin. If the intercept deviates from this expectation, publication bias may be the cause. This would occur, for instance, when small studies are disproportionately associated with larger effect sizes.

Another method of assessing publication bias is the "fail safe N" (Rosenthal, 1979), which estimates the number of nonsignificant studies needed to reduce the overall ES to nonsignificance. This may be too conservative a procedure, however, since missing studies would rarely have an effect size of zero. Orwin's (1983) method employs the same idea, but calculates the number of studies with a specific effect size (not necessarily 0 as in Rosenthal's method) needed to reduce the overall effect to

whatever value the researcher designates as clinical nonsignificance and thus will result in lower values than Rosenthal's method.

Independence

When proceeding with a meta-analysis, the researcher must be certain that effect sizes are independent of each other. Many studies report outcomes using more than one measurement instrument. For example, dietary fat can be measured by self report, by calculation from dietary recall, by percent calories from fat, or percent of people reporting attainment of the Action or Maintenance stages of the TTM. Commonality across studies, reliability, and validity must be considered when choosing the one measure or they may be averaged. Additionally, various measures can be compared across studies to determine if one may under- or over-estimate the effect.

Outliers

One weakness commonly associated with employing a mean as an outcome is that an inordinately large or small effect size can skew the result. Since such outliers can arise from mis-coded data or the occasional odd finding they should be examined and the coding checked to insure accuracy of data. The analyst can keep them and move them to the closest cluster (Lipsey & Wilson, 2001) or employ sample-adjusted meta-analytic deviancy statistic (Huffcutt & Arthur, 1995). Unfortunately this uses rather subjective scree plots, involves numerous computations, and is likely to remove small correlations.

Modeling Variance

Once the data has been cleaned and effect sizes appropriately adjusted, analysis can proceed. As in any other statistical procedure, employing a mean with a large variance does not provide a precise representation of the population value. In meta-analysis, then, the variance among effect sizes carries prime importance and is known as homogeneity testing.

Homogeneity of Effect Size

Just as in any other statistical procedure, the method of modeling the variance affects procedures, assumptions, drawbacks, and conclusions. In meta-analysis, the heterogeneity among studies is the variance in question. Before pooling estimates we need to see if they can reasonably be described as sharing a common effect size. In other words, we perform a 'null hypothesis test' on the assumption that error is due to sampling error or systematic variance. Meta-analyses employs Hedges Q (a Chi-square with $df = k-1$) for this test. A significant result suggests heterogeneity and a presence of moderators. The homogeneity analysis is calculated using the equation: $Q = (\sum \omega_i ES_i^2) - [(\sum \omega_i ES_i)^2 / \sum \omega_i]$, where $\omega_i = [2(n_1 n_2) (n_1 + n_2 - 2)] / (n_1 + n_2) [t^2 + 2(n_1 + n_2 - 2)]$. Unfortunately this test has low power when assumptions of normality are not met and when variances are not equal. It fails to reject the null even with large differences, yielding false models and false pooling of variance estimates (Harwell, 1997). Harwell (1997) found that it does, however, work well when study sample sizes are proportionally greater than the number of studies included (k).

The variance among studies can be modeled in three ways: fixed effects, random effects, or a combination of both, referred to as a mixed model. The results of

the Q test have been used to suggest which model to employ, but some researchers disagree with this determination. Rosenthal (1995) suggests that contrasts should be planned and done independent of the heterogeneity test. He states, "A significant X^2 for heterogeneity 'morally' obligates one to search for moderators, but a nonsignificant X^2 does not preclude the search." Lipsey and Wilson (2001) suggest a significant Q test is enough to determine model used. Hedges & Vevea (1998) suggest that choice of model depends primarily upon the *nature of the inference* desired. Heterogeneity is not the sole criterion for choosing a model. Fixed and random effect models have different inherent assumptions and techniques that affect the inference drawn from them.

Fixed Effects Modeling

Fixed effects modeling treats variability between studies as random error resulting from subject-level sampling error. Hedges & Vevea (1998) call the fixed effect model the "conditionally random effects" model because it allows inferences conditional upon only the sample of effect sizes at hand. It assumes the effect sizes are a complete sample and creates a mean effect size without statistical modeling. The fixed effects model has high type I error rates (up to .50) because it underestimates variances (Cohn & Becker, 2003; Overton, 1998) and is not conservative.

Random Effects Modeling

The random effects model treats variability between studies as sampling error plus a randomly distributed other source of variability ("study-level" error). It assumes the effect sizes at hand are randomly drawn from a population of studies and thus estimates a population mean effect size from a sampling distribution. It thus allows

‘unconditional’ inferences beyond the observed studies. A difference between fixed and random methods is seen only when studies are very heterogeneous. This procedure may overestimate variances, leading to more conservative estimates with wider confidence intervals than with fixed effects (Overton, 1998).

Mixed Effect Models

The ANOVA analog groups effect sizes of descriptive variables such as gender into ‘between’ and ‘within’ categories and tests homogeneity using Chi-square within and between groups. If significant residuals result from these tests, an additional component random effect can be assumed to exist, resulting in a mixed effect model. The mixed model has lower Type I error than a fixed model, but less power for detecting moderators. A sensitivity analysis can be done to compare the fixed and random models.

Meta-Regression

With continuous variables a weighted multiple regression can also be done to explain heterogeneity. This procedure has high type I error rates for detecting moderators when a large amount of heterogeneity is present. When employing regression techniques, correlations should be examined between descriptive variables to assess for colinearity. Macros have been written in Stata, SPSS, and SAS that permit regression with corrections for standard errors unique to meta-analysis. These programs output an overall fit statistic, Q_R for the regression and a Q_E for the residual error, which are distributed as a chi-square (Lipsey & Wilson, 2001). They also output an overall R^2 for the model allowing examination of variance accounted for and change in R^2 when adding additional predictors.

Choice of Model

Choice of model then is a statistical and theoretical decision. Hunter and Schmidt (2000), for example, conclude that random effects modeling should usually be used because it allows generalization, while Lipsey and Wilson (2001) suggest the model is difficult to estimate. Meta-analysis is thus not immune to statistical problems. Just as in other statistical procedures, meta-analysis requires decisions that affect conclusions drawn from the analyses.

With these characteristics in mind, the analyst will still probably proceed as Lipsey and Wilson (2001) suggest, by doing the Q test, and if significant, (1) assuming random effects, (2) assuming excess variance is not random, accepting a fixed-effect model with post hoc tests, or (3) assuming a mixed effects model such that that error beyond subject level error is both systematic *and* random.

The overall effect size significance test depends on choice of model. The fixed effect model employs the Stouffer method where all Z's are added and divided by k or the lower confidence limit method (L. V. Hedges, Cooper, & Bushman, 1992). This method usually agrees with Stouffer but has higher type I error. Random effect models use a one sample t-test on the mean effect size, but as discussed, are more conservative than fixed effects procedures with higher type II error.

Effect Size Interpretation

In interpreting the meaning of an effect, use of standardized effect sizes facilitates analysis with commonly understood indices. For continuous outcomes using Hedges g as the effect size measure, g is directly comparable to a Z-score and interpreted as a normal distribution with a mean of 0 and a standard deviation of 1. An

effect of $g = .30$, for example, indicates that the intervention group is 1/3 of a standard deviation above the control group and exceeds the scores of 62% of the control group. A normal distribution is assumed, however, in this example as well as outcomes presented with the same measure. Interpretation of a standardized effect involves calculating the mean and pooling the standard deviation of the control groups of the included studies. This provides a baseline from which to compare the effect size of the intervention.

Interpretation of dichotomous outcomes using the odds ratio is more common in the literature and therefore more readily understood. The odds ratio measures the relative effect of the treatment group versus the control group. Thus an OR of 1.30 represents a 30% greater effect over the control group. Again, knowledge of the control group mean is necessary for translation to the original metric.

Results of a meta-analysis can be interpreted using Cohen's (1988) suggestions of effect sizes of small = .20, medium = .50, and large = .80 ($r = .10, .25, .40$, or 1%, 6%, and 14% of variance accounted for). Actuarial studies have somewhat supported his claim, defining small effects as $d < .30$, medium as $d = .50$, and large as $d > .67$. These are general guidelines and all effect sizes should be interpreted in light of the content area. For example, public health interventions account for .05%, 1.0% and 1.5% of variance for small, medium, and large effect sizes respectively (Rossi, 2003). The mean effect size, however, can be misleading without an examination of amount and sources of variation in the effect sizes contributing to those means. The analyst must explore moderators and sample size before being confident in the estimate. Results can be translated then into other metrics such as original measurement metric

by determining its mean and deviation, and into a Binomial Effect Size Display. The BESD shows correlations of effect sizes in terms of overlapping distributions. Another comparison is the criterion contrast, a comparison of the effect size with a known difference of practical significance. For example, the effect size could be compared to a 5% difference in smoking cessation rates usually considered to be clinically significant.

Power

One benefit of meta-analysis is the ability to estimate a population parameter estimate from under-powered studies. Power for each study can be calculated to quantify the number of studies that have clinically significant effect sizes compared to the population mean, yet would be considered nonsignificant due to low sample size. Power for detecting significance of the overall mean effect size has not been considered an important issue in meta-analysis since the technique is less interested in statistical tests than in obtaining population estimates. Power calculations can be done using Cohen's tables with the obtained effect size as the estimate. Of greater importance for meta-analysis is determining the power of the Q test for heterogeneity since this can indicate the presence of moderators and choice of statistical model. It has been suggested that when sample size is below 10, the Q test has limited power to detect heterogeneity (Lipsey & Wilson, 2001).

Confidence Intervals

Confidence intervals are useful in illustrating the precision of individual effect sizes and the overall parameter estimate. Displaying intervals for each effect size making up the mean enables quick examination of the point estimates, error variance,

and statistical significance of each study. For the parameter estimate, their width relates to amount of data, level of confidence chosen, and the model employed. Fixed effect CI's may tend to be smaller than those from a random effects model.

Limitations of Meta-Analysis

Despite the many applications and broad conclusions that can be drawn from meta-analysis, the procedure has various drawbacks. While acceptable, meta-analyses are correct regarding direction of effect about 80% of the time (Naylor, 1997). First, the procedure is relatively new and lacks refined techniques. Techniques basic to statistics such as ANOVA and multiple regression cannot be run with common software packages without advanced knowledge. Statistically, meta-analysis places emphasis on the variances from individual studies. Variance challenges the assumption that the studies really do measure the same construct, and also affects the Q test (Harwell, 1997). Lipsey and Wilson (2001) assert that analysts need to determine the source of this variance by using analyses of methodology. Rosenthal's "coefficient of robustness" (Rosenthal, 1995) can be used to weight means by their variability. Difficulty in using common statistical programs with meta-analysis limits use of multivariate techniques and possibly more accurate, specific conclusions. In addition meta-analyses may not have the sample sizes required to perform multivariate analyses.

The process of meta-analysis can also bias results. The old computer adage "garbage in - garbage out" applies to meta-analysis as well. The effect size estimate is only as good as the studies that compose it. If studies use a limited sample, generalizability will be limited. For example, a disproportionate number of

psychotherapy studies are done using Cognitive-Behavioral Therapy techniques, biasing results of meta-analyses to CBT over other forms of therapy. Poor design techniques and lack of control will result in effect sizes that fail to validly sample reality. Meta-analysts recommend investigating if results differ according to study methodological quality. Multiple regression models can be used in this determination in which methodological features predict effect size, with the beta weights indicating the influence of each factor. Confounding of substantive and methodological features also occurs. If a difference appears in two groups that are also measured differently, we cannot determine the source of the discrepancy (Kazdin & Weisz, 1998).

Bias inevitably enters a meta-analysis from publication bias as well. Meta-analysts may not be able to locate a certain kind of study or fail to search properly. Even with a good search, the field has documented the publication bias problem such that significant studies are more often published than non-significant studies. This results in upward bias of the mean effect size. Since a systematic effort will locate published and non-published studies, the analyst can compare effect sizes for published and unpublished studies. The fail safe N, regression methods, funnel plots and imputational trim and fill techniques provide multiple methods of estimating and correcting the effect size for publication bias. A different form of publication bias can also enter when various authors use data from the same study, resulting in multiple inclusion of the same effect size (Naylor, 1997).

Meta-analysis in Sum

The research community has created a problem by relying on significance testing without question, resulting in an overemphasis on replication. Too much

information exists without the ability to gain knowledge from it. Meta-analysis has been developed to answer this current crisis in research. By compiling many similar studies of important questions, meta-analysis allows treatment effect estimates in terms of both direction and magnitude. Despite its drawbacks, which for the most part can be mitigated, meta-analysis is becoming the procedure of choice for compiling results and for informing policy. As it becomes the accepted standard, it may help focus research on issues necessary to move research forward: power, sample size, effect size, and confidence intervals. It will help to solve past controversies, and as databases grow, suggest moderators that may better inform future interventions. Meta-analysis provides hope for moving past the information age into an age of cumulative, constructive knowledge.

CHAPTER 3: RESEARCH JUSTIFICATION AND PREDICTIONS

Meta-analysis and Tailored Interventions

Previous reviews of ‘first generation’ tailoring studies (Brug, Campbell, & van Assema, 1999; Skinner et al., 1999; Strecher, 1999) summarized significant findings from published articles. Kroeze, Werkman and Brug (2006) conducted a systematic review of computer tailored interventions for physical activity and dietary behaviors. While they conducted a thorough literature search, they concluded that “heterogeneity of the included studies hindered the pooling of data.” This method, while a first step in evaluating an emerging literature, provides little insight into the strength of intervention effects and makes comparison among studies difficult if not impossible, especially given differing sample sizes. To deepen our understanding of the key variables involved in effective tailoring, studies must be compared systematically using effect sizes and moderators.

In previous reviews meta-analytic methods could not be used because targeted behaviors, tailoring methods, and populations differed widely among studies (Skinner et al., 1999). Since then the number of tailored interventions has increased dramatically facilitating the use of meta-analytic methods. The present study will also broaden its scope beyond previous reviews that concentrated only on smoking (Strecher, 1999) or nutrition (Brug et al., 1999) to include a full range of tailored interventions. Including studies focusing on smoking, nutrition, physical activity, mammography, sexual behavior, and alcohol use will increase the sample size of the data and permit comparisons on key variables common across studies. Using moderators in the analysis enables disentangling of method, substance, and error

(Marsh et al., 2001). Multiple regression models will be used to determine if methodological features predict effect size, beta weights indicating influence of each factor. The technique can also compare the methodological quality of studies to determine if various conditions affected the outcomes (Lipsey & Wilson, 2001).

Research Hypotheses and Predictions

This research will investigate the following predictions:

- Within the same behaviors (i.e. smoking cessation, dietary fat reduction) tailored interventions will outperform non or minimally tailored interventions.
- Interventions employing iterative feedback (re-tailored reports) will outperform reports tailored with one data collection and feedback timepoint (Brug et al., 1998; Lutz et al., 1999; Velicer et al., 1999).
- Effect size estimates will increase with outcome assessment timepoints (Campbell et al., 2002; Kristal et al., 2000; Prochaska & Velicer, 2004).
- Proactive recruitment strategies will result in a small percentage of participants reaching behavioral criteria but will reach a larger percentage of people than reactive methods (Prochaska & Velicer, 2004).
- Number of variables intervened upon (i.e. stage of change, decisional balance, self-efficacy) will increase the effectiveness of main outcomes (Anderson et al., 2001; Oenema et al., 2005; Lutz et al., 1999).
- Theoretical orientation employed will not influence outcome, since the main behavior change theories (i.e. TTM, HBM, TRA) share similar constructs and have all demonstrated support (Noar & Zimmerman, 2005).

- Study group/site will not influence effect size estimates (Noar & Zimmerman, 2005)..
- Moderators such as ethnic background, stage of change, amount of smoking (light/heavy), decisional balance, etc. will affect treatment outcomes.

CHAPTER 4: METHODS

Literature Searches

The computerized databases PsychInfo and Medline will be searched for relevant studies during the spring and summer of 2006. Additionally, reference lists from published studies and personal communications with authors will be used to locate studies. The effort to broaden the search beyond published studies helps to limit publication bias for significant results showing large effects. Datasets will include published articles, conference presentations, and papers in progress.

Inclusion Criteria

Databases will be searched starting in 1988 (the year of the first tailored feedback study). Studies must have employed a tailored intervention, have included a comparison group, and must have given participants feedback reports, whether printed or computer-based. An intervention will be considered “tailored” if it provides individual-based feedback on at least one assessed variable. Studies will need to contain information regarding sample size, outcome variables, means and standard deviations for treatment effects and/or test statistics.

Coding

In the current meta-analysis, each behavior in a multiple behavior study will be looked at separately. Studies will be coded according to the coding scheme as outlined in Appendix A. Given the nature of the project, the author will read and code all studies. To enhance accuracy, coding will be re-examined after a delay of a few

months prior to data analysis. Studies will be entered using the Comprehensive Meta-Analysis software package.

Effect Size Calculation

For outcomes measured in continuous format such as minutes of physical activity per week or servings of fruit per day, Hedges *g* will be used to calculate effect size. This method has received the most support for its accuracy in determining effect sizes. Hedges *g* is a derivation of the mean difference (*d*) effect size. Cohen's *d* is simply the mean difference divided by the pooled standard deviation of the two groups defined by the equation:

$$d = \frac{mean_1 - mean_2}{\sqrt{(SD_1^2 + SD_2^2)/2}} \quad (2.)$$

Cohen's *d* will not be used in the present study because it does not account for sample size, nor unequal sample size between groups, causing *d* to be biased in the direction of the larger standard deviation (and the less reliable effect). The mean difference divides by the simple additive standard deviation of each mean, whereas *g* corrects for sample size bias by dividing by a denominator corrected for sample size (*n*-1), thereby correcting for underestimation of population standard deviation. Hedges' *g* requires means, SD's, and N's for each group and is defined as the difference between the sample means divided by the average pooled sample standard deviation as shown below in equation 3 (Hedges & Olkin, 1985). In addition, Hedges found that *g* can be upwardly biased when samples sizes are less than 20 per group. The second part of the following equation provides this correction.

$$g = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1 - 1)SD_1^2 + (n_2 - 1)SD_2^2}{N_{total} - 2}}} \times \left(1 - \frac{3}{4(n_1 + n_2) - 9}\right) \quad (3.)$$

This equation also illustrates that with a large N little difference will exist between estimates of g and d .

In the present study many outcomes will be presented in terms of proportion of the sample attaining various behavioral criteria such as percent reaching Action or Maintenance stages of change, Action traditionally being defined as engaging in the desired behavior, but for less than six months, and Maintenance being defined as sustaining the behavior change for more than six months (Velicer, et al., 2000). Standardized mean difference effect size indices do not directly apply in these instances. Effect size for proportional outcomes, therefore, will be calculated using odds ratios. The odds ratio is defined as

$$OR = \frac{a/b}{c/d} \quad (4.)$$

where

	Status A	Not Status A
Group 1	a	b
Group 2	c	d

Unfortunately, effect size cannot directly be calculated from odds ratio as the format is not standardized. Taking the natural log of the odds ratio, the log odds, thereby

standardizes the odds ratio on a scale from -1 to 1, with a mean of 0. This transformation facilitates standardized comparisons above and below the mean, unlike the odds ratio where, for example, an OR of -0.5 below the mean is equivalent to an OR of 2 above the mean. The resulting log odds can then be converted back to the more easily interpreted odds ratio.

When combining results across studies, outcomes often will be reported in either continuous or dichotomous outcome formats. This presents a difficulty in choosing a combined effect size measure due to lack of equivalency between odds ratio and standardized mean difference. Outcomes can be calculated separately for each index but this results in a decreased number of studies available for comparison. Transformations of the odd ratio into standardized mean difference effect sizes are available. For analyses in which both exist, but many are reported in dichotomous format, Lipsey and Wilson (2001) suggest the logit transformation of the odds ratio enabling reporting of effects as a standardized mean difference. This method will be employed for dietary intake and exercise outcomes when results are reported in both continuous (e.g. number of fruits and vegetables/day) and dichotomous outcome formats (e.g. % reaching Action or Maintenance stages for fruit and vegetable intake).

Weighting of Studies

The main benefit of meta-analysis is the ability to pool effects from a variety of small N studies to arrive at an overall estimate of effect size. Since the analysis will likely include studies with relatively small samples and others with large samples, simply taking the arithmetic mean of effects does not account for accuracy in estimating population means. It is assumed that larger samples will result in more

accurate estimates of effect size and therefore should receive greater weight in the combination of effects. Prior to estimation of overall effect, each obtained effect size will be weighted by the inverse variance weight. Effects employing Hedges g will be multiplied by their weight (ω) prior to combination and weighted according to equations 5, 6, and 7:

$$ES' = \left[1 - \left(\frac{3}{4N - 9} \right) \right] ES \quad (5.)$$

$$SE = \sqrt{\frac{n_1 + n_2}{n_1 n_2} + \frac{(ES')^2}{2(n_1 + n_2)}} \quad (6.)$$

$$\omega = \frac{1}{SE^2} = \frac{2n_1 n_2 (n_1 + n_2)}{2(n_1 + n_2)^2 + n_1 n_2 (ES')^2} \quad (7.)$$

where ES = the observed (uncorrected) effect size estimate, ES' = the corrected effect size, N = the total sample size, SE = the standard error of the corrected effect size estimate, n_1 and n_2 = the sample sizes of the two groups, and ω = the inverse variance weight. Odds ratios will first be transformed to log odds and weighted according to equations 8, 9, and 10:

$$ES_{LOR} = \log_e (ES_{OR}) \quad (8.)$$

$$SE_{LOR} = \sqrt{\left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d} \right)} \quad (9.)$$

$$\omega_{LOR} = \frac{1}{SE_{LOR}^2} = \frac{abcd}{ab(c + d) + cd(a + b)} \quad (10.)$$

Outliers

As in any statistical analysis, outliers can unduly influence outcomes where the arithmetic mean is used to combine effects. Such outliers can result from miscoding, the presence of moderators, publication bias, or the occasional odd finding. In meta-analysis the presence of outliers differentially effects the estimation of the fixed and random effects models as outlined below. In the present study when effects fall outside two or three standard deviations of the overall mean, they will be examined for accuracy of coding. The analysis program for the present study permits analysis of the overall effect with “one study removed.” This analysis will be followed when outliers are present to assess their effect on the mean. Studies with small sample size may have little effect on the overall estimate given that they are first weighted. To preserve as much data as possible, studies will be deleted from analysis only if the extreme value indicates that the study does not conceptually fit with the set of comparisons studied.

Confidence Intervals

The point value of a mean effects size for a group of studies is considered the best estimate of the overall effect, but given the standard error, the actual population effect could fall within a range of values. Calculating a confidence interval around each effect accounts for the standard error of the estimate and permits simple analysis of the range in which the population effect may lie for a certain level of confidence. Confidence intervals for each study and overall mean effect will be calculated according to equations 11, 12, and 13:

$$SE_{ES} = \sqrt{\frac{1}{\sum \omega_i}} \quad (11.)$$

$$ES_L = ES - z_{(1-\alpha)}(SE_{ES}) \quad (12.)$$

$$ES_U = ES + z_{(1-\alpha)}(SE_{ES}) \quad (13.)$$

where SE_{ES} is the standard error of the effect size mean, ω_i is the inverse weight associated with the effect size i with $i = 1$ to k effect sizes included in the mean, ES is the mean effect size, and $z_{(1-\alpha)}$ is the critical value for the z-distribution (1.96 for $\alpha = .05$). Therefore, if the 95% confidence interval is chosen, the standard error is multiplied by the corresponding Z-value (1.96) and one can be 95% confident that the population value falls in this range. The confidence interval will become larger if the 99% level is chosen and smaller if the 90% level is chosen. Additionally, the interval is affected by the precision of the estimate, estimated by the standard error. Larger studies will offer more precise estimates with tighter intervals. This method also permits analysis of significance such that if the value includes 0, the null is maintained at the chosen significance level. Confidence intervals are suited to graphical display, which facilitates visual analysis of the range of effects and their precision and will be reported in table and graphical formats.

Modeling Variance

As outlined previously, the pattern of variability of effect sizes around the mean is of prime concern in meta-analysis. The variation of effect sizes can be assumed to occur from sampling error among the subset of studies, from systematic variation, or from a combination of sampling and random error. Various methods of modeling can account for each of these instances.

Fixed effects modeling assumes that the only source of variance in the sample of effects arises from the actual variability of the sample of effects around the mean. Fixed effects assumes that the mean value represents the best value of the population of scores and that variance around this mean arises from subject-level error alone. This variability will be tested with the Q test done on the fixed effects variance component according to equation 14:

$$Q = (\sum \omega_i ES_i^2) - \frac{(\sum \omega_i ES_i)^2}{\sum \omega_i} \quad (14.)$$

where ES_i is the individual effect size for i number of effects and ω_i is the weight for each effect. Q is distributed and interpreted as a Chi-square test. A significant Q test indicates that additional variance beyond that expected for the given N exists in the scores. Significance may indicate the presence of moderators such as age, recruitment method, or intervention strategy.

Given that the Q test has low power to detect differences with less than 10 scores, planned moderator analysis will be done in the present study on all means. The following demographic and theoretical moderators will be tested: mean sample age, percent female, percent minority, retention rate, recruitment strategy (proactive v. reactive), intervention strategy (tailored v. retailored), and study group. Dichotomous moderators can be tested using procedures similar to the ANOVA, or regression analysis can be used for analysis of discrete and continuous moderators simultaneously. The ANOVA for meta-analysis partitions variance using the same techniques as any other ANOVA by separating the total variance Q (Total) into Q (Between) and Q (Within) using the equations:

$$Q_b = (\sum \omega_j \overline{ES}_j^2) - \frac{(\sum \omega_j \overline{ES}_j)^2}{\sum \omega_j} \quad (15.)$$

where Q_b is the between groups variance, $\overline{ES}_j$ is the weighted mean effect size for each group, ω_j is the sum of the weights within each group, and j is the number of groups.

$$Q_w = \sum \omega_i (ES_i - \overline{ES}_j)^2 \quad (16.)$$

where Q_w is the pooled Q within groups variance, ES_i is the individual effect size, $\overline{ES}_j$ is the weighted mean effect size for each group, ω_j is the sum of the weights within each group, i is the number of effect sizes, and j is the number of groups. The between groups Q is the measure of interest and is tested with the Chi-square distribution. A mixed effects analysis can also be employed. This model tests variance left over after assuming a within groups random effects model. Since it accounts for more within-groups variance, this model therefore has less statistical power to detect between-groups effects than the fixed effects model. The ANOVA or mixed effects models will be employed when only discrete predictors are of interest. Meta-regression will be employed when both discrete and continuous variables require investigation. Correlations among variables will be examined to indicate possible inclusion in the regression. Common regression procedures do not correctly estimate standard errors and statistical test values for effect sizes, and thus, require correction. Lipsey and Wilson have written a SPSS macro, which will be used in the present study, that performs corrected meta-regression (Lipsey & Wilson, 2001).

As in any statistical test, detection of moderators using significance testing becomes difficult with small N's. With a sample size of 10 studies (as for many comparisons in the present study), power for detecting differences between groups is minimal, ranging from .07 with an ES of $d = .20$ to .55 with an (unlikely) ES of $d = 1.0$, assuming $\alpha = .05$. Thus moderators may be present, but will not be able to be detected statistically. When moderators are not found in the sample of studies, the random effects variance component will be assumed for theoretical purposes. A random effects model is preferred in this instance because generalization to a larger population of studies is desired and because populations are assumed to have pre-existing differences, and studies are assumed to randomly vary in characteristics such as sampling strategy, recruitment, message content, etc. Accounting for this non-systematic error theoretically permits generalization of the mean effect to the larger population of similar studies. The random effects model may provide a slightly different estimate of the mean effect since fixed effects modeling weights smaller studies less than random effects modeling. The weights assigned to random effects are more balanced across small and large studies. That is, a random effects model operates from the assumption that extreme values, whether from large or small studies, come from a population of values and thus give small and large studies a more equal weighting than fixed effects. With a few small but extreme effects, such as in the case of publication bias, the random effects mean may be upwardly biased, but will have a larger variance given that it includes between-study variance. Random effects modeling includes the addition of a random variance component (ν_θ) to the fixed effects subject-level error (ν_i) and is thus defined as:

$$\nu_i^* = \nu_\theta + \nu_i \quad (17.)$$

where ν_θ is the random or between-subjects component, and ν_i is the subject-level sampling error. Defining the random component presents a difficulty and can be accomplished using the method of moments or maximum likelihood estimation. Since it is iterative, maximum likelihood can provide slightly more accurate estimates, but its difficulty outweighs the difference and the methods of moments will be employed in the present study defining ν_θ as:

$$\nu_\theta = \frac{Q - (k - 1)}{\sum \omega_i - \left(\frac{\sum \omega_i^2}{\sum \omega_i} \right)} \quad (18.)$$

where Q is the value of the homogeneity test, k is the number of effect sizes and ω_i is the inverse weight for each effect size.

Statistically Dependent Effect Sizes

In the present study many instances exist in which studies include one control group with multiple intervention groups, all measured at more than one timepoint. Unfortunately, inclusion of more than one comparison or outcome timepoint per subgroup introduces statistical dependence for which usual statistical procedures do not account. Gleser and Olkin (1994) have developed methods of accounting for such covariance among outcomes, thus enabling inclusion of otherwise lost data. Their method, however, requires knowledge of the correlation between outcome measures, which is rarely provided. Therefore, the present study will follow the suggestion of Lipsey and Wilson (2001) and combine outcomes where possible. The mean of

timepoints will be used and outcomes will also be grouped for separate comparison. Statistical procedures cannot be carried out in this instance, but it will permit examination of overall trends.

Choice of Comparison

A true test of tailoring requires comparison of tailored studies not only to no-intervention control group, but to minimal or usual intervention whether that be providing targeted pamphlets, informational brochures, or physician advice. In the present study when studies include comparison of tailored intervention with assessment only control and with minimal intervention, minimal intervention comparison will be chosen as the reference group for effect size calculation. Assessment-only reference groups will be combined with minimal intervention in combining studies to increase the overall N. This is theoretically feasible given that assessment itself is well-known to introduce intervention effects. Various combinations and definitions of tailoring were discovered during data coding. In the present analysis studies that provided at least one assessment and feedback will be referred to as “tailored.” Studies completing an assessment and feedback at more than one timepoint will be considered “retailored.” Studies that assessed participants only once, yet provided feedback on more than one occasion are termed “multiple tailored” and were grouped for analysis with the tailored studies. When tailored and retailored modalities will be considered together, the term “re/tailored” will be employed.

Missing Data Procedures

Missing data has long been neglected in many outcome analyses and can result in biased inferences. Sophisticated iterative techniques exist for analysis of subject

data, but do not apply well to meta-analysis given the differentiation among studies and small sample sizes. Data in meta-analysis can be missing at the study level (studies unable to be located), at the effect size level (studies that do not report sufficient data for calculation) and at the moderator level (lack of reporting characteristic variables). Study-level missingness in meta-analysis is dealt with in terms of publication bias analyses. Effect size missingness can be dealt with by exclusion of a study or inclusion of a best predicted value. If a study did not supply enough information for calculating effect size it was not included in the analysis. If a study indicated that the effect was nonsignificant but did not specify group differences, it was included with the ES entered as 0 along with a dummy variable indicating this. Such a procedure can downwardly bias the result and can even be counterintuitive to the overall rationale for meta-analysis. Thus inclusion of a dummy variable allows effect of the study on the overall mean to be easily indicated and assessed. Methods of dealing with missing data common to all data analysis include complete case analysis, substitution of the mean, and analysis of available data. In the present analysis of moderator variables complete case analysis will be preferred for analysis of moderators, but given that not all studies report data required for the present moderators, available case analysis will then be chosen if enough studies remain to enable a comparison. Given the variability among studies, mean substitution will not be employed as it is unlikely to give an accurate estimate of the missing value.

Type of analysis carried out in each particular study also bears relevance to missing data. In cases where studies report results from both intent-to-treat and all subject available conditions, effect sizes from all subjects available will be used for

analysis. As Hall et al. (2001) show, intent-to-treat analyses make the erroneous and unnecessary assumption that subjects who drop out should be considered unsuccessful in terms of an intervention, thereby underestimating effect sizes.

Mean effects will be assessed for degree of publication bias using four techniques: (1) Rosenthal's fail-safe N, (2) Orwin's fail safe N, (3) Egger's regression intercept, and (4) Duval and Tweedie's trim and fill technique. Rosenthal's fail-safe N calculates the number of studies with an effect size of 0 needed to reduce the overall effect to statistical nonsignificance (usually defined as $p > .05$) and will be calculated as:

$$N_{fs} = \left(\frac{N_o}{Z_c^2} \right) (N_o \bar{Z}_o^2 - Z_c^2) \quad (19.)$$

where N is the number of studies, Z_c is the critical value of Z, and $\bar{Z}_o$ is the overall mean effect size. Rosenthal's method has been criticized, however, due to its reliance on statistical, versus clinical significance. Orwin (1983) modified the idea to account for level of clinical significance and the assumption that missing studies would have a mean ES of 0:

$$N_{fs} = \frac{N(\bar{d} - d_c)}{d_c - d_{fs}} \quad (20.)$$

where N is the number of studies, $\bar{d}$ is the average ES, and d_c is the criterion value such that d would equal if N_{fs} number of studies with a mean ES of d_{fs} were included. Thus criteria for clinical nonsignificance can be included. In the present study the mean ES for nonincluded studies will be set to 0 (to facilitate comparison to Rosenthal's) and the level of clinical nonsignificance set to $g = .10$ or $OR = 1.18$

which would represent a difference of about 6% between treatment and control, the minimum for clinical significance in population-based interventions.

In the Egger test, the standardized effect (effect size divided by standard error) is regressed on precision (inverse of standard error). A significant intercept suggests that bias is present in the studies such that treatment effect is related to precision of estimation (study quality).

Trim and fill is a technique developed by Duval and Tweedie (2000) and assesses the symmetry of the funnel plot. Small studies with negative effects are not likely to get published and would fall on the bottom left of the funnel plot. When publication bias exists, a disproportionate number of studies will fall to the bottom right of the plot. This technique determines the number of asymmetrical studies, imputes their counterparts to the left, and calculates a new mean effect size. Each study is ranked using the equation:

$$r_i^* = \text{rank}(|d_i - \bar{d}|) \quad (21.)$$

where r_i^* is the rank for each study ES and d_i is the study effect and $\bar{d}$ is the mean effect. R_0 is the imputed number of studies where:

$$R_0 = \gamma^* - 1 = 0 - 1 = -1 \quad (22.)$$

$$\gamma^* = k - r_h^* = 10 - 10 = 0 \quad (23.)$$

where k is the total number of studies and r_h^* is the largest negative rank. They suggest that publication bias exists when $R_0 > 3$. Publication bias is not expected to be of significant concern in the present meta-analysis as most tailored interventions arise

from grant-funded research, carrying with it the obligation to publish results, regardless of significance.

CHAPTER 5: RESULTS

General Characteristics of the Studies

The dates of inclusion for the current analysis range from 1988 through August 2006. The literature search began with entering the most general search term “tailor*” into the Medline and Psychlit databases, the “*” indicating retrieval of derivations of tailor such as tailored, tailoring, etc. Additionally, the search terms “computer*” and “tailor*” and “expert system” were entered as well. The general search terms were used so as to perform an inclusive database search. Manual examination of titles and abstracts was deemed necessary to find as many relevant studies as possible given that studies could have varying titles. These search terms resulted in retrieval of 958 references. Each was examined by the author and 126 were selected for full-text retrieval. Most studies were obtained from full-text databases and through interlibrary loan. Five studies were obtained from author correspondence. Examination of references from obtained studies and review articles resulted in the addition of 81 articles and studies for a total of 207 relevant papers. Three were dissertations that could not be obtained through library loan. Results from interventions were reported by 135 studies, but of 13 unique behaviors intervened upon, only four contained sufficient numbers of effect sizes to include for analysis – smoking cessation, dietary behavior, physical activity, and mammography screening (see Table 1). Thus 74 studies intervening on these four behaviors were included for analysis, representing 96,018 participants (see Table 2.). Studies were coded according to a coding scheme modified from one previously developed to code theory-based behavior change studies (Hall, 2005). Fifty-three variables were coded which accounted for over 30,000 unique

pieces of data (see Appendix A for the list of variables coded). All but one study had been published in a peer-reviewed journal. Five authors were contacted for additional information regarding outcomes and additional studies. Three responded resulting in the inclusion of two studies. Another author preferred not to have the data included until publication.

Table 1: Number of Studies Located by Behavior

Behavior	# Included	# Excluded	Total # Located
Alcohol Use		2*	2
Diet	16	18	34
Diet Smoking	1		1
Diet Mammography	1		1
Diet Physical Activity	6	2	8
Diet Smoking Sun Protection	1		1
Diet Smoking Sun Mammography	1		1
Diet, Physical Activity, Smoking Mammography, Cancer screening, Physical activity, Seat belt use	1		1
Diet Physical Activity Smoking	2		2
Injury Prevention		2*	2
Mammography	10	9	19
Organ Donation		1*	1
Pain Management		1*	1
Risk Reduction		2*	
Physical Activity	11	8	19
Cancer Screening		1*	1

Behavior	# Included	# Excluded	Total # Located
Sexual Risk Prevention		2*	2
Smoking	24	14	38
Stress Reduction		1*	1
Total	74	63	135

* Studies reporting usable outcome, but excluded due to inadequate sample size for meta-analysis.

Table 2: Studies Included for Analysis

Study	Behaviors	Outcomes	N	Intervention Methods	Recruitment	Theory
Anderson, Winett, Wojcik, Winett, & Bowden, 2001	Diet	- % Cal Fat - Fiber g/day - Fruit and Veg/day	464	Computerized assessment and tailored fdbk at store kiosk v no tx delivered at grocery store	Proactive. Recruited in person at grocery store	SCT
Armitage & Conner, 2001	Diet	- % Cal Fat - Fat Score	801	Brochure v behavior feedback	Proactive. Employees contacted at worksite.	na
Aveyard, Griffin, Lawrence, & Cheng, 2003	Smoking	7 Day Quit	2471	Multiple intervention comparison between brochure, manual, counseling, and tailored feedback	Proactive. Random sample recruited through clinic	TTM
Bastani, Maxwell, Bradford, Das, & Yan, 1999	Mammog	% getting mammog	902	Brochure v behavior feedback + manual	Proactive. RDD	na
Bock, Marcus, Pinto, & Forsyth, 2001	Phys Act	% A or M 7 PAR	194	Brochure v 4 tailored letters and manual delivered at home	Reactive. Fliers and announcements	TTM SCT
Borland, Balmford, & Hunt, 2004	Smoking	7 Day Quit 6 mo Quit	1058	Brochure v 3 Tailored letters delivered at home	Reactive. Callers to cancer info line	na
Borland, Balmford, Segar, Livingston, & Owen, 2003	Smoking	7 Day Quit	1578	Manual, 3 tailored, counseling calls delivered at home	Reactive. Callers to cancer info line.	TTM
Brug, van Assema, & de Vries, 1996	Diet	- Fat Score - Fruit/day - Veg/day	507	Brochure v 1 tailored letter	Proactive. Worksite	SCT
Brug, Glanz, Van Assema, Kok, & Van Breukelen, 1998	Diet	- Fat score - Fruit/day - Veg/day	762	Comparison between brochure, tailored, and tailored delivered at home	Reactive. Fliers and announcements	na
Brug, Steenhuis, van Assema, Glanz, & De Vries, 1999	Diet	- Fat score - Fruit/day - Veg/day	347	Feedback v 1 tailored delivered at home	Proactive. Worksite sample	na

Study	Behaviors	Outcomes	N	Intervention Methods	Recruitment	Theory
Bull, Jamrozik, & Blanksby, 1999	Physical Activity	% increasing activity	763	Assessment, brochure, tailored, and physician advice delivered at home and clinic	Proactive. Clinic sample	TTM SCT
Bull, Kreuter, & Scharff, 1999	Physical Activity	% increasing activity	272	Assessment, tailored, brochure, and personalized delivered at home	Proactive. Clinic sample	TTM
Burnett, Magel, Harrington, & Taylor, 1989	Diet	- Fat Score - Fiber/day	77	Brochure v 1 tailored delivered at school	Proactive. High school students.	na
Campbell et al., 1999	Diet	- Fruit and Veg/day	459	Assessment, tailored, culturally tailored delivered at home	Proactive. Church members	SCT
Campbell et al., 2004	Diet	- Fat Score	526	Assessment or tailored delivered at clinic via computer	Proactive. Clinic members requested to participate	TTM SCT
Campbell, DeVellis, Strecher, & Ammerman, 1994	Diet	- Fat Score - Fruit and Veg/day	558	Assessment, brochure, or tailored letter delivered at home	Proactive. Clinic members requested to participate	TTM SCT
Campbell et al., 2002	Diet, Phys Act Smoking	- Fat Score - Fruit and Veg/day - 7 DPAR - 7 day quit	859	Assessment or retailored letter delivered at home	Proactive. Worksite requested to participate	TTM SCT
Campbell et al., 2004	Diet	- Fat Score - Fruit and Veg/day	410	Assessment or 1 tailored interactive computer feedback delivered at clinic	Proactive. Recruited through clinic	TTM SCT
Cardinal & Sachs, 1995	Phys Act	% Stage Progress	113	Behavior fdbk or tailored letter delivered at home	Recruited through worksite	TTM
Champion et al., 2002	Mammog	% getting mammog	1367	Assessment, counseling calls, or tailored letter delivered at home	Proactive. Random sample of HMO members	TTM HBM
Champion et al., 2007	Mammog	% getting mammog	1245	Assessment, counseling calls, or tailored letter delivered at home	Proactive. HMO members	TTM HBM

Study	Behaviors	Outcomes	N	Intervention Methods	Recruitment	Theory
Clark et al., 2002	Mammog	% getting mammog	1324	Assessment, brochure, or retailedd letter delivered at home	Proactive. HMO members	TTM
Curry, Wagner, & Grothaus, 1991	Smoking	24 hr quit 7 day quit 6 mo quit	1217	Brochure or retailedd letter delivered at home	Reactive. Respondents to fliers and announcements	SCT
Curry, McBride, Grothaus, Louie, & Wagner, 1995	Smoking	7 day quit 6 mo quit	1317	Assessment, manual, tailored+manual, tailored+manual+call, letters and calls delivered at home	Proactive. RDD sample	SCT
Delichatsios et al., 2001	Diet Phys Act	- % Cal Fat - Fiber g/day - Fruit/day	298	Automated calls delivered at home	Proactive. HMO members	SCT
Dijkstra, De Vries, & Roijackers, 1999	Smoking	7 day quit	843	Assessment, manual, 1 tailored and 3 tailored letters delivered at home	Reactive. Recruited through fliers and announcements	TTM
Dijkstra, 2005	Smoking	24 hr quit	202	Brochor or personalized feedback delivered through interactive computer at University setting	Proactive. Recruited through school	na
Elder et al., 2005	Diet	- % Cal fat - Fat score - Fiber g/day	357	Targeted brochure, tailored, or tailored+coiunseling delivered at home	Proactive. RDD	na
Etter & Perneger, 2004	Smoking	28 day quit 7 day quit	2934	Assessment, brochure or 4 retailedd letters delivered at home	Proactive. Random sample contacted by mail	TTM
Etter, 2005	Smoking	7 day quit	11969	Brochure or tailored delivered through computer over Internet	Reactive. Respondents to smoking cessation website.	TTM
Gould & Anderson, 2000	Diet	Fat score	95	Counselor or interactive computer tailored delivered at clinic	Proactive. Recruited from clinic patients	TTM

Study	Behaviors	Outcomes	N	Intervention Methods	Recruitment	Theory
Greaney et al., 2007	Phys Act	7 DPAR % A/M % stage progress	1274	Assessment or retailored+manual+counselor calls and letters delivered at home	Reactive. Respondents to fliers and announcements.	TTM
Heimendinger et al., 2005	Diet	Fruit and veg/day	3402	Brochure, 1 tailored, 4 tailored, or retailored letters delivered at home.	Proactive. Recruited through cancer info line	TTM HBM SCT
Jacobs et al., 2004	Diet Phys Act	- % A/M - Fat score - % increasing phys act	511	Advice or 6 tailored letters+counselor call delivered at home	Reactive. Recruited from clinic patients	TTM SCT
Johnson, Driskell, Johnson, Dymont et al., 2006	Diet Phys Act Med adherence	% A/M	404	Assessment or retailored letters + manual	Proactive. RDD	TTM
Jones & Rossi, 2003	Diet Smoking	% A/M	1029	Advice or multiple tailored+counselor calls and letters delivered at home	Proactive. Random sample of HMO members	TTM
King et al., 2006	Phys Act	7 day PAR	335	Behavior fdbk or Print feedback + motivational interview delivered in clinic	Proactive. HMO members	SCT
Kosma, Cardinal, & McCubbin, 2005	Phys Act	- % stage progress - PAR	151	Assessment or 4 tailored interactive website interventions	Reactive. Respondents to fliers and announcements	TTM
Kreuter & Strecher, 1996	Mammog Smoking Diet Cancer Screening Phys Act Seat Belt Use	- % getting mammog - 7 day quit - fat score - % CDC criteria	1317	Assessment, blvr fdbk, or 1 tailored letter delivered at home	Proactive. Recruited through clinic patients	TTM

Study	Behaviors	Outcomes	N	Intervention Methods	Recruitment	Theory
Kreuter et al., 2005	Diet Mammog	- % getting mammog - Fruit and veg/day	720	Assessment, tailored or culturally tailored letters delivered at home	Proactive. Recruited through clinic patients	na
Kristal, Curry, Shattuck, Feng, & Li, 2000	Diet	- Fruit and veg/day - % A/M - Fat score	1459	Assessment or tailored letter + manual + counselor calls delivered at home	Proactive. Recruited through HMO members	TTM SCT
Lennox et al., 2001	Smoking	28d quit	2553	Assessment, personalized or tailored letter delivered at home	Proactive. Recruited from clinic patients	TTM
Lipkus, Lyna, & Rimer, 1999	Smoking	28d quit	266	Advice, tailored letter + advice, tailored letter + advice + counselor call	Proactive. Recruited from random sample of clinic patients	TTM
Lipkus, Rimer, Halabi, & Strigo, 2000	Mammog	% getting mammog	1934	Reminder, tailored letter, or counselor call delivered at home	Proactive. Recruited from sample of HMO members	TTM
Lutz et al., 1999	Diet	Fruit and veg/day	710	Assessment, brochure, or tailored letter delivered at home	Proactive. Recruited from random sample of HMO members	TTM SCT HBM
Marcus et al., 1998a	Phys Act	% stage progress	1559	Brochure or 2 tailored letters delivered at home	Reactive. Recruited from worksite sample	TTM
Napolitano et al., 2003	Phys Act	- % stage progress - 7d PAR	65	Assessment or 4 tailored interactive web-based sessions delivered at home	Reactive. Recruited from worksite	TTM SCT
Oenema, Tan, & Brug, 2005	Diet	- Fat score - Fruit/day - Veg/day	782	Assessment, brochure, or 2 tailored feedback sessions delivered by CD- ROM at home	Proactive. Recruited from worksite	PAPM

Study	Behaviors	Outcomes	N	Intervention Methods	Recruitment	Theory
Owen, Ewins, & Lee, 1989	Smoking	7d quit	208	Assessment, brochure, or tailored letters delivered at home	Reactive. Respondents to fliers/announcements	na
Patrick et al., 2001	Phys Act Diet	-7d PAR -Fruit and veg/day -Fat score	148	Computerized tailored feedback or tailored feedback + follow up letters or calls	Proactive. Recruited from high school students	RP
Peterson & Aldana, 1999	Phys Act	% stage progress	784	Assessment, brochure or tailored letter delivered at worksite	Proactive. Recruited from worksite	TTM
Pinto et al., 2002	Phys Act	% CDC criteria 7d PAR	298	Assessment or 6 tailored counselor calls delivered at home	Proactive. Recruited from HMO members	TTM SCT
Prochaska, Velicer, Fava, Rossi, & Tsoh, 2001	Smoking	- 24h quit - 7d quit - 28d quit - 6m quit	4144	Assessment or tailored letters delivered at home	Proactive. RDD sample	TTM
Prochaska, DiClemente, Velicer, & Rossi, 1993	Smoking	- 24h quit - 6m quit	756	Brochure, manual, tailored letter or tailored + manual + counselor call conditions delivered at home	Reactive. Recruited from respondents to fliers/announcements	TTM
Prochaska et al., 2001	Smoking	- 24h quit - 7d quit	1447	Assessment, tailored letters, tailored letters + manual, or tailored letter + manual + counselor call delivered at home	Proactive. Recruited from HMO members	TTM
Prochaska et al., 2004	Smoking Diet Sun protection	- % A/M - 24h quit - 7d quit	2460	Assessment or tailored letters + manual delivered at home	Proactive. Recruited from parents of participants in another study	TTM
Prochaska et al., 2005	Smoking Diet Sun protection Mammog	- % A/M - Fruit and veg/day - 24h quit	5407	Assessment or tailored letters + manual delivered at home	Proactive. Recruited from clinic patients	TTM

Study	Behaviors	Outcomes	N	Intervention Methods	Recruitment	Theory
Prochaska & Sallis, 2004	Phys Act	7d PAR	138	Assessment or tailored feedback provided via computer in school	Proactive. Recruited from middle school student population	na
Rakowski et al., 1998	Mammog	% getting mammog	1864	Assessment, brochure, or 2 tailored letters delivered at home	Proactive. Recruited from random sample of HMO members	TTM
Rakowski et al., 2003	Mammog	% getting mammog	2023	Reminder letter or tailored letter delivered at home at either 2 mos or 10 mos post-assessment	Proactive. Recruited from sample of HMO members	TTM
Rimer, Orleans, Fleisher, & Cristinzio, 1994	Smoking	7d quit	1867	Brochure, manual, or manual + counselor calls	Reactive. Recruited from respondents to fliers/announcements	TTM
Rimer et al., 2002	Mammog	% getting mammog	1287	Reminder letter, tailored letter, or tailored letter + counselor calls	Proactive. Recruited from HMO members	TTM PAPM
Saywell, Champion, Skinner, Menon, & Daggy, 2004	Mammog	% getting mammog	1390	Assessment, tailored letter, or tailored letter + counselor calls delivered at home	Proactive. Recruited from HMO members	HBM
Shiffman, Paty, Rohay, Di Marino & Gitchell, 2000	Smoking	- 28d quit - 10wk quit	3627	Brochure + audiotape + nicotine gum or tailored letter + gum or tailored letter, gum, and counselor call	Reactive. Recruited from calls to quit line included with nicotine gum	RP
Shiffman, Paty, Rohay, Di Marino, & Gitchell, 2001	Smoking	- 28d quit - 10m quit	4209	Manual+nicotine patch or tailored print feedback + nicotine patch delivered at home	Reactive. Recruited from callers to quit line	RP
Skinner, Strecher, & Hospers, 1994	Mammog	% getting mammog	489	Brochure or tailored letter delivered at home	Proactive. Recruited from clinic patients	TTM HBM

Study	Behaviors	Outcomes	N	Intervention Methods	Recruitment	Theory
Stevens, Glasgow, Toobert, Karanja, & Smith, 2003	Diet	- % Cal Fat - Fat score - Fruit and veg/day	616	Assessment or tailored letter + manual + counselor calls delivered at kiosk and at home	Proactive. Recruited from HMO members	na
Strecher et al., 2005	Smoking	- 28d quit	3971	Manual + nicotine patch or tailored print feedback + nicotine patch delivered at home	Reactive. Recruited from callers to info line	na
Strecher, Kreuter, Den Boer, & Kobrin, 1994	Smoking	7d quit	296	Assessment or tailored letter delivered at clinic	Proactive. Recruited from clinic patients	TTM HBM ATT
Strecher, Shiffman, & West, 2005	Smoking	7d quit	72	Brochure, tailored, multiple tailored, or retailed letters delivered at home	Reactive. Recruited from callers to info line	TTM HBM ATT
Vandelanotte, De Bourdeaudhuij, Sallis, Spittaels, & Brug, 2005	Diet Phys Act	% Cal Fat Fat score % CDC phys act	1023	Assessment or retailed computer feedback + manual delivered at university lab	Reactive. Recruited from respondents to fliers/announcements	TTM TPB
Velicer, Prochaska, Fava, Laforge, & Rossi, 1999	Smoking	7d quit	2882	Manual, tailored, or 2, 3, or 6 retailed letters delivered at home	Proactive. Recruited from HMO members	TTM
Velicer et al., 2006	Smoking	24hr quit 7d quit 6m quit	2054	Manual, tailored, or tailored letters + call in combination with patch delivered at home	Proactive. Recruited from VA hospital population	TTM

TTM = Transtheoretical Model, HBM = Health Belief Model, TPB = Theory of Planned Behavior, PAM = Precaution Adoption Process Model, RP = Relapse Prevention, ATT = Attribution theory, SCT = Social Cognitive Theory

Table 3 reports general characteristics of the studies. A total of 21 variables were reported to have been intervened upon and a total of seven different health behavior change theories were employed. A few studies reportedly drew from more than one theory, resulting in 10 different combinations of theories. Table 4 summarizes health behavior theories referred to in the studies by actual variables intervened upon. Studies reporting use of the TTM accounted for 60% of interventions, with an additional nine studies employing stage of change without mention of the TTM, resulting in 75% of studies mentioning stage of change. Use of stage of change is often erroneously equated with the TTM model, without inclusion of three additional components – decisional balance, self efficacy, and processes of change - that developers of the TTM view as essential to the model and its effectiveness. Thus, this meta-analysis analyzed the frequency of use of each of the TTM components, results of which are presented in Table 5.

Table 3: General Characteristics of the Studies

	k	%	Mean(SD)	Median
Recruitment Strategy				
Proactive	52	69.3		
Reactive	22	29.3		
NA	1	1.3		
Random Sampling	18	24.0		
Delivery Site				
Home	62	82.7		
Home + Clinic	1	1.3		
Clinic	5	6.7		
Kiosk	2	2.7		
School	2	2.7		
University Lab	2	2.7		
Worksite	2	2.7		
Intervention Method				
Counselor Calls	4	4.7		
Automated Calls	2	2.4		
Print + Counselor Calls	14	16.7		
Print Alone	53	63.1		
Interactive Computer	11	13.1		
Recruitment Strategy				
HMO	18	24.3		
Clinic	16	21.3		
Worksite	9	11.5		
Call in Center	8	10.8		
RDD	5	6.8		
School	4	5.4		
Church	1	1.4		
Random Mailing	1	1.4		
Store	1	1.4		
Website	1	1.4		
Country				
US	58	78.7		
Non-US	16	21.3		
Behaviors Intervened Upon				
One	63	85.1		
Two	7	9.5		
Three	2	2.7		
Four	2	2.7		
Recruitment Rate (%)	47		59.1 (33)	66.9
Retention Rate (%)	74		75.3 (14.9)	77.3
Mean Age	68		41.9 (11.4)	40.6
% Female	74		71.1 (21.2)	66.4
% Minority	67		23.5 (29.4)	10.0

Table 4: Reported Health Change Theory by Theoretical Variables Employed for Feedback

	Total	Not Rptd	TTM	HBM	SCT	RP	TTM+ PAM	TTM+ TPB	TTM+ SCT	TTM+ HBM	TTM+ HBM+ ATT	TTM+ HBM+ SCT
Stage of Change	57	4	32	1	2	0	1	1	8	5	1	2
Decisional Balance	32	1	22	1	1	0	1	0	2	2	0	2
Self-Efficacy	44	6	21	2	5	2	0	1	3	2	0	2
Processes	24	1	22	0	0	0	0	0	1	0	0	0
Strategies	14	3	5	0	1	3	0	0	2	0	0	0
Social Support	10	2	2	0	2	1	0	1	2	0	0	0
Motives	10	3	3	0	2	1	0	0	1	0	0	0
Benefits	6	0	0	1	0	1	0	0	1	2	1	0
Barriers	25	4	7	2	2	1	1	1	3	3	1	2
Affect	1	0	0	0	0	0	0	0	1	0	0	0
Attitudes	4	2	2	0	0	0	0	0	0	0	0	0
Attributions	1	0	0	0	1	0	0	0	0	0	0	0
Health Beliefs	4	2	0	0	0	0	0	0	0	1	0	1
Bhvr Fdbk	27	9	5	1	5	0	1	1	4	1	0	0
Goal Setting	12	2	1	0	2	1	0	0	4	0	0	2
Knowledge	4	1	0	0	0	0	0	0	2	1	0	0
Risk	12	2	1	2	1	0	1	0	1	4	0	0
Culture	5	2	3	0	0	0	0	0	0	0	0	0
Symptoms	1	1	0	0	0	0	0	0	0	0	0	0
Addiction Level	5	2	2	0	0	1	0	0	0	0	0	0
Negative Views	1	1	0	0	0	0	0	0	0	0	0	0
Total	13	32	2	7	3	1	1.3	1	8	4	1	2
% of Total	17.3	42.7	2.7	9.3	4.0	1.3	1.3	1.3	10.7	6.7	1.3	2.7

**Table 5: Studies Mentioning Use of TTM and Comparison of TTM Components
Used in Tailored Interventions**

Component and Matching Studies	k	%
Stage Only	10	21%
Bull & Kreuter et al 1999		
Campbell et al 2002		
Campbell et al 2004		
Jacos et al 2004		
Kristal et al 2000		
Lipkus et al 2000		
Marcus et al 1998		
Napolitano et al 2003		
Skinner et al 1994		
Strecher et al 1994		
Stage + Decisional Bal	3	6%
Campbell et al 1994		
Lipkus et al 1999		
Rimer et al 2002		
Stage + Self Efficacy	5	11%
Campbell & Honess et al 1994		
Champion et al 2006		
Rimer et al 1994		
Strecher et al 2005		
Vandelanotte et al 2005		
Stage + Processes	1	2%
Peterson et al 1999		
Stage + Dec Bal + Processes	4	9%
Cardinal et al 1995		
Clark et al 2002		
Rakowski et al 1998		
Rakowski et al 2003		
Stage + Decisional Bal + Self Efficacy	5	11%
Bull & Jamrozik et al 1999		
Champion et al 2002		
Etter et al 2004		
Heimendinger et al 2005		
Lutz et al 1999		
Stage + Self Eff + Processes	1	2%
Etter et al 2005		

Component and Matching Studies	k	%
Full TTM	18	38%
Aveyard et al 2003		
Bock et al 2001		
Borland et al 2003		
Dijkstra et al 1999		
Gould et al 2000		
Greaney et al 2007		
Jones et al 2003		
Kosma et al 2005		
Lennox et al 2001		
Pinto et al 2002		
Prochaska et al 2001		
Prochaska et al 1993		
Prochaska et al 2001a		
Prochaska et al 2004		
Prochaska et al 2005		
S.Johnson et al 2006		
Velicer et al 1999		
Velicer et al 2006		
Total	47	100%

Mammography

For mammography screening behavior, 12 studies were found that employed a print tailoring intervention component. Comparisons that were similar across studies were combined, and four studies were chosen as the fewest to analyze together. With these restrictions, four overall combinations of comparisons were extracted. The most common comparison involved assessment or minimal intervention, such as a general brochure, versus print tailoring. Eleven studies were included in this analysis, with outcome timepoints ranging from two to 24 months post baseline. Most studies reported results as proportion of participants obtaining mammography using odds ratios and thus log odds will be employed for effect size analyses.

Results by Comparison

Assessment or Minimal Intervention versus Print Tailored

Eleven studies were included in the largest comparison group for mammography screening. Included in this combined comparison were studies that compared assessment only or standardized brochure with theory-based tailored print feedback. Only two studies employed one additional retailored feedback that presented subjects with progress made since their first assessment and thus were included with the other nine. Mean effect size for the eleven studies with fixed and random effects was $LO = .22 (.04)$, $Z = 5.58$, $p = .001$ and $LO = .24 (.07)$, $Z = 5.58$, $p = .001$. The test for heterogeneity reached significance where $Q = 22.71$, $p = .012$, $df = 10$ (see Table 6 and Table 7). All studies employed proactive recruitment from a non-treatment seeking population and thus recruitment strategy did not serve as a moderator. No significant differences were found by study group, by location of recruitment (HMO

membership, RDD, clinic, etc), or by age group. Meta-regression showed that studies with greater retention rates showed significantly larger effect sizes ($B = -1.41, (.60), p = .018$). Inclusion of more ethnic minorities in studies also predicted increased effect sizes over and above retention rate ($B = .44 (.21), p = .036$). Rosenthal's fail safe N for these comparisons was 50, indicating confidence that the effect size from these studies is representative of this mode of intervention. Orwin's fail safe N was 16 using $LO = 1.09$ as the critical value for clinical significance. Both examination of the funnel plot from these studies and the results of Egger's regression procedure ($I = .29 (.86), t = .255$) suggest minimal publication bias within these comparisons. Trim and fill did not suggest any imputed values.

Assessment or Minimal Intervention versus Print Tailored + Tailored Calls

Four studies employed tailored feedback in addition to one or two phone conversations also based on tailored feedback. The mean ES for this intervention versus assessment or minimal intervention was $LO = .63 (.10), Z = 6.39, p = .0001$ with both fixed and random effects models (see). These studies were homogenous where $Q = 2.5, p = .475, df = 3$. The small number of studies in this comparison prevents the search for moderators. Results of Rosenthal's (38) and Orwin's fail safe N's (28) and Eggers regression ($I = -8.86 (5.93), t = 1.49, p = .14$) indicated little effect of publication bias.

Assessment or Minimal Intervention versus Tailored Calls

Four studies included a telephone-only condition in which tailored feedback was provided in the context of one or two phone contacts. The mean ES for this comparison was $LO = .37 (.19), p = .0001$ under both models (see Table 6 and Table

8). Test of heterogeneity was nonsignificant where $Q = 1.79, p = .62, df = 3$. The small number of studies in this comparison prevents the search for moderators. Fail safe N for these comparisons was 15 and 14 with Rosenthal's and Orwin's methods, respectively. Eggers regression intercept ($I = 11.61 (3.17), t = 3.66, p = .02$) was significant, indicating that this comparison may be subject to publication bias given that all studies showed relatively large effects.

Print Tailored versus Print Tailored + Tailored Calls

Four studies included both a print tailored and a print tailored plus tailored call condition, enabling comparison on the additive effect of phone contact over print feedback alone. The mean ES for this comparison was $LO = .26 (.09), Z = 2.929, p = .003$ under both models (see Table 6 and Table 8). Standard error for these studies was small and thus the Q test for heterogeneity was nonsignificant where $Q = .95, p = .814, df = 3$. The small number of studies in this comparison prevents the search for moderators. Fail safe N for these comparisons was 5 and 9 with Rosenthal's and Orwin's methods, respectively, yet Eggers regression intercept was nonsignificant ($I = .824 (2.12), t = .39, p = .37$), suggesting that the effect size is small, yet comprised of homogenous studies indicating some confidence in the ES of this comparison.

Table 6: Mammography Mean ES

Comparison	k	Model	LO	95% CI	Z	p	Q	P (df)	Fail Safe N R/O
Assess or minimal v. re/tailored	11	Fixed Random	.22 .24	.14 -.30 .09 -.38	5.68 3.15	.001 .002	22.71	.012(10)	50/16
Assess or minimal v. tailored + calls	4	Fixed Random	.63 .63	.44 -.82 .44 -.82	6.39 6.39	.0001 .0001	2.49	.47(3)	38/28
Assess or minimal v. calls	4	Fixed Random	.37 .37	.19 -.55 .19 -.55	4.03 4.03	.0001	1.79	.62(3)	14/15
Tailored v. Tailored + calls	4	Fixed Random	.26 .26	.09 -.43 .09 -.43	2.93 2.93	.003 .003	.95	.81(3)	5/9

Effects over Time

Figure 1 presents the summary of effects across time for the 11 re/tailored and four studies that added counselor calls to the intervention. Data is taken from studies assessing outcomes at more than one timepoint and thus dependence prohibits statistical comparison. The trend reveals decreases across time for both modalities, with the call condition remaining superior. Examination of confidence intervals suggests no significant differences between the groups, however, and a significant drop in effectiveness after 6-month assessment.

Figure 1: Effect of Re/Tailored and Re/Tailored + Calls on Mammography

Screening

Table 7: Minimal Intervention v. Re/Tailored on Percent Getting Mammography

Study	Comparison	Outcome	Time point	LO	SE	Z	p
Bastani 1999	Brochure v Feedback+Man	% mammography	12	0.37	0.17	2.21	0.03
Champion 2002	Assess v tailored	% mammography	2	0.54	0.19	2.80	0.01
Champion 2006	Assess v tailored	% mammography	4	0.54	0.20	2.72	0.01
Clark 2002	Assess v retailored	% mammography	14	0.21	0.05	3.95	0.00
Kreuter & Strecher 1996	Assess v tailored	% mammography	6	-0.04	0.55	-0.07	0.94
Kreuter & Sugg 2005	Assess v tailored	% mammography	18	0.42	0.41	1.04	0.30
Lipkus 2000	Reminder v tailored	% mammography	12, 24	0.05	0.16	0.34	0.74
Rakowski 1998	Assess v retailored	% mammography	18	0.36	0.13	2.68	0.01
Rakowski 2003	Reminder v 10 mo tailored	% mammography	15	0.03	0.16	0.16	0.87
Rimer 2002	Reminder v tailored	% mammography	12	-0.36	0.18	-2.02	0.04
Saywell 2004	Assess v tailored	% mammography	2	0.45	0.18	2.59	0.01
Fixed				0.22	0.04	5.68	0.00
Random				0.24	0.07	3.15	0.00

Table 8: Effect of Counselor Calls Percent Getting Mammography

Study	Comparison	Outcome	Time point	LO	SE	Z	p
Champion 2002	Assess v tailored + call	% mammography	2	0.77	0.20	3.86	0.00
Champion 2006	Assess v tailored + call	% mammography	4	0.65	0.20	3.30	0.00
Rimer 2002	reminder v tailored + call	% mammography	12	0.34	0.22	1.56	0.12
Saywell 2004	Assess v tailored + call	% mammography	2	0.70	0.18	3.89	0.00
Fixed				0.63	0.10	6.39	0.00
Random				0.63	0.10	6.39	0.00
Champion 2002	Assess v call	% mammography	2	0.51	0.20	2.52	0.01
Champion 2006	Assess v call	% mammography	4	0.47	0.20	2.31	0.02
Lipkus 2000	Reminder v call	% mammography	12,24	0.20	0.17	1.20	0.23
Saywell 2004	Assess v call	% mammography	2	0.40	0.18	2.20	0.03
Fixed				0.37	0.09	4.02	0.00
Random				0.37	0.09	4.02	0.00
Champion 2002	Tailored v Tailored + call	% mammography	2	0.36	0.24	1.52	0.13
Champion 2006	Tailored v Tailored + call	% mammography	4	0.13	0.17	0.80	0.42
Rimer 2002	Tailored v Tailored + call	% mammography	15	0.33	0.15	2.15	0.03
Saywell 2004	Tailored v Tailored+MI	% mammography	2	0.25	0.18	1.39	0.16
Fixed				0.26	0.09	2.93	0.00
Random				0.26	0.09	2.93	0.00

Diet

For healthy eating behavior, 27 studies were found that employed a print tailoring intervention component. Of these studies, eight intervened solely on decreasing dietary fat and four intervened solely on increasing fruit and vegetable intake. Fifteen studies intervened on two or more diet components such as both fat and fruit and vegetable intake. Dietary outcomes will be analyzed separately.

Results by Comparison

Fat intake

No differences were found between outcomes measuring fat intake – percent calories from fat, food frequency fat scores, and percent reaching Action or Maintenance stages for fat intake. Percent of calories from fat showed the greatest standard error (.033), followed by percentage Action or Maintenance (.026), and fat scores (.022). Percent calories from fat relies on two calculations based on self-report – total calories and fat intake – creating possibility for additional error variance. Fat scores will be the preferred outcome variable since it is measured continuously, followed by percent reaching Action or Maintenance stages, followed by percent calories from fat if necessary. Hedges g will be employed as the effect size measure for these studies since most studies reported data in continuous format. Dichotomous outcomes, when they occurred, were transformed with the logit function.

Fat Intake – Assessment or Minimal Intervention versus Tailored and Retailored

Studies employing either tailored or retailored interventions were included together. This allows the combination of 17 effect sizes. Employing a random effects model, the mean effect size for these studies was $g = .20$ (.03), $z = 8.86$, $p = .0001$ (see

Table 9 and Table 10). These studies showed nonsignificant heterogeneity where $\nu = .004$, $Q(16) = 23.24$, $p = .11$. Coding studies by tailored or retailored intervention strategy revealed significant between groups differences where $Q_b(1) = 4.38$, $p = .04$. Within groups variance was reduced to nonsignificance where $Q_w(15) = 18.87$, $p = .22$, suggesting that intervention strategy accounted for some of the variability among studies. Rosenthal's fail safe N for these studies was 298 and Orwin's was 17 (with $g = .10$ as the criterion), indicating confidence that a representative sample of studies is included. Egger's regression showed a nonsignificant finding suggesting minimal effects of publication bias ($I = .45 (.67)$, $t = .67$, $df = 16$, $p = .26$). Trim and fill suggested the imputation of four values to the left of the mean, but this would only decrease the overall ES slightly to .19.

Fat Intake - Assessment or Minimal Intervention versus Tailored

Ten studies employing an assessment or minimal intervention compared to tailored feedback were compared first. The overall effect size employing the fixed effects model was $g = .15 (.03)$, $p = .0001$ (see Table 9 and Table 10). The overall test for heterogeneity was significant where $\nu = .009$; $Q = 16.12 (df = 9)$, $p = .06$. The effect size employing the random effects model showed $g = .16 (.05)$, $p = .0001$. Examination of the funnel plot revealed two studies with effect sizes above .70. Both of these studies had large standard errors (.29 and .37), and thus represent small N studies included possibly due to publication bias. Employing the trim and fill technique for imputation of studies, the fixed effects mean was reduced to $g = .14$. If these two studies were removed from the analysis, the ES = .15 (.03), matching the estimated effect after bias correction. This resulted in a nonsignificant test for

heterogeneity where $\nu = .006$, $Q (df = 7) = 2.18$, $p = .10$. Only one study employed a reactive recruitment strategy, four studies were conducted in the Netherlands, three at the University of North Carolina and one in the UK. Results did not differ among these groups $Q (2) = 3.86$, $p = .15$. In terms of continuous predictors, percent female alone showed a significant relationship with the effect size $B = -.26 (.13)$, $Q (1) = 3.98$, $p = .04$, with a nonsignificant residual $Q (9) = 13.82$, $p = .13$. See Figure 2 (larger points indicating studies with larger N's). Fail safe N for the eight studies was 79 and 8 with Rosenthal's and Orwin's estimations, respectively, and Egger's regression was nonsignificant suggesting minimal effects of publication bias ($I = .96$ (1.05), $t = .91$, $df = 9$, $p = .19$). Trim and fill suggested imputation of 2 values to the left of the mean, reducing the effect to $g = .16$.

Figure 2: Regression of Percent Female on Hedges g: Dietary Fat

Fat Intake - Assessment or Minimal Intervention versus Multiple or Retailored

Seven studies compared assessment only or minimal intervention to multiple tailored or iteratively retailored interventions. These studies were examined together for a mean effect size of $g = .25 (.03)$, $p = .001$. Effect sizes ranged from .18 to .46

with no outliers present. Studies were homogenous where $\nu = .002$, $Q(6) = 2.71$, $p = .84$ (see Table 9 and Table 10). Moderators were not present. Examination of the fail-safe N for these studies with Rosenthal's calculation was 90 and 10 with Orwin's. Egger's regression ($I = .68 (.55)$, $t = 1.18$, $df = 5$, $p = .15$) suggests little effect of publication bias. Trim and fill imputed two values to the left of the mean, reducing the overall effect size minimally to .24.

Fat Intake - Assessment or Minimal Intervention versus Tailored+ Calls

Four studies compared brief advice or assessment only conditions to tailored feedback plus one or more brief telephone contacts. Effect sizes ranged from .06 to .50. The mean effect size for these studies was $g = .25 (.07)$, $p = .0001$ using the random effects model. Tests were found to be heterogeneous where $\nu = .012$, $Q(4) = 9.87$, $p = .04$ (see Table 9). One study was removed (Jones & Rossi, 2003) because it showed an effect twice as much as other studies and also employed a differing outcome measure (% Action or Maintenance). Removing this study reduced the heterogeneity to nonsignificance where $\nu = .000$, $Q(3) = 2.8$, $p = .42$. Removal of this study resulted in a mean effect size of $g = .22 (.04)$, $p = .0001$ under both models. Sample size prevented search for moderators. Duval and Tweedie's trim and fill suggested no need for imputed values. Rosenthal's fail safe N was 21 and Orwin's was 5 for these studies with Egger's regression ($I = -.55 (1.80)$, $t = .30$, $df = 2$, $p = .40$) nonsignificant, suggesting minimal publication bias.

Table 9: Dietary Fat Mean ES

Comparison	k	Model	g	SE	95% CI	Z	p	Q	p (df)	Fail Safe N R/O
Assess or minimal v. Re/tailored	17	Fixed Random	.20 .22	.03 .03	.15-.24 .17-.27	8.86 8.32	.0001 .0001	23.24	.10(16)	298/1
Tailored	10	Fixed Random	.15 .16	.03 .05	.09-.21 .07-.25	4.80 3.42	.0001 .0001	16.13	.06(9)	
Retailored	7	Fixed Random	.25 .25	.03 .03	.18-.31 .18-.31	7.73 7.74	.0001 .0001	2.73	.84(6)	
Assess or minimal v. Tailored + calls	4	Fixed Random	.22 .22	.04 .04	.14 -.30 .14 -.30	5.23 5.23	.0001 .0001	2.80	.42 (3)	21/6

Fat Intake – Effects over Time

Effects increased over time for the tailored and retailored interventions, with an increasing, albeit fairly unreliable, effect even at 13-24 outcomes. With the addition of counselor calls to the print component, effects were initially higher ($g = .28$) but decreased across time ($g = .18$), becoming smaller than the print tailored conditions. Even though formal analysis could not be done, a horizontal trendline can be drawn incorporating all confidence intervals, suggesting that no significant differences exist across timepoints, nor between groups. See Figure 3.

Figure 3: Re/Tailored and Re/Tailored + Calls for Dietary Fat Reduction

Table 10: Minimal Intervention v. Tailored and Retailored Intervention for Fat Intake

Study name	Comparison	Outcome	Time point	g	SE	Z	P
Anderson et al 2001	Assess v Retailored	% Calories From Fat	6	0.26	0.16	1.65	0.10
Campbell et al 2002	Assess v Retailored	FFQ Fat Score	6	0.20	0.08	2.58	0.01
Delichatsios et al 2001	Exercise v Retailored	% Calories From Fat	6	0.37	0.19	1.96	0.05
Prochaska et al 2004	Assess v Retailored+Manl	% A or M Fat	12,24	0.18	0.07	2.56	0.01
Prochaska et al 2005	Assess v Retailored+Manl	% A or M Fat	12,24	0.27	0.05	5.42	0.00
S.Johnson et al 2006	Assess v Retailored+Manl	% A or M Fat	18	0.46	0.23	2.00	0.05
Vandelanotte et al 2005	Assess v Retailored+Manl	FFQ Fat Score	6	0.27	0.09	3.09	0.00
Retailored Fixed							
Retailored Random							
Armitage & Conner 2001	Brochure v Bhvr Feedback	FFQ Fat Score	5	0.23	0.09	2.58	0.01
Brug et al 1996	Brochure v Tailored	FFQ Fat Score	1	0.28	0.11	2.58	0.01
Brug et al 1998	Brochure v Tailored	FFQ Fat Score	2	0.24	0.10	2.49	0.01
Brug et al 1999	Bhvr Feedback v Tailored	FFQ Fat Score	1	0.13	0.06	2.33	0.02
Burnett et al 1989	Brochure v Tailored	FFQ Fat Score	3	0.71	0.29	2.46	0.01
Campbell & Honess et al 1994	Assess v Tailored	FFQ Fat Score	3	0.03	0.10	0.33	0.74
Campbell et al 1994	Assess v Tailored	FFQ Fat Score	4	0.27	0.12	2.13	0.03
Campbell et al 2004	Assess v Tailored	FFQ Fat Score	2	-0.15	0.11	-1.30	0.19
Kreuter & Strecher 1996	Bhvr Feedback v Tailored	FFQ Fat Score	6	0.32	0.42	0.77	0.44
Oenema et al 2005	Assess v Multiple Tailored	FFQ Fat Score	1	0.11	0.10	1.13	0.26
Tailored Fixed							
Tailored Random							
Overall Fixed							
Overall Random

Fruit and Vegetable Intake – Assessment or Minimal Intervention versus Tailored and Retailored

Nine studies employed tailored or retailored intervention to increase fruit and vegetable intake together. For the fixed model the mean effect was $g = .18 (.02)$, $p = .0001$ and $g = .20 (.03)$, $p = .0001$ with random effects. Heterogeneity was present where $\nu = .0007$, $Q(8) = 16.48$, $p = .04$. Comparing the tailored versus retailored methods revealed that the five studies employing a retailored intervention showed mean effect of $g = .22(.03)$, $p = .0001$ and the six studies employing a tailored intervention had a mean effect size of $g = .11 (.04)$, $p = .0001$, which was a significant difference $Q_b(1) = 4.82$, $p = .03$ (see Table 11 and Table 12). Heterogeneity was still present in the tailored studies $Q(5) = 11.65$, $p = .04$, and percent female was the only significant predictor where $B = -.73 (.37)$, $p = .05$ with nonsignificant residual $Q(4) = 6.48$, $p = .16$. See Figure 4. As a group, Duval and Tweedie's trim and fill technique suggested no imputed values. Fail safe N's were 66 and 8 and the results of Egger's regression ($I = -.99 (.88)$, $t = 1.24$, $df = 7$, $p = .15$) indicate possible, but negligible effects from publication bias. The effect of the tailored and retailored interventions increased up to 6-month assessment. Only three studies presented outcomes after 12 months, one of which was the group outlier and negative. Figure 5 presents the long-term effects with and without the outlier removed. Confidence intervals suggest no differences across timepoint, possibly due to small sample size, especially at 1-6 months.

Figure 4: Regression of Percent Female on Hedges's g: Fruit and Vegetable Intake

Figure 5: Effect of Re/Tailored Intervention on Fr/Veg Intake Over Time

Fruit Intake – Assessment or Minimal Intervention versus Tailored and Retailored

Four studies intervened on fruit and vegetable intake but measured each food category separately. One other study intervened on fruit intake only. These five studies were examined as a group. For fruit outcome, the mean effect size was $g = .17 (.05)$, $p = .001$. Effect sizes ranged from $g = .01$ to $.32$, yet heterogeneity was not present where $\nu = .0001$, $Q(4) = 3.92$, $p = .42$. See Table 11 and Table 13. A trend was present for the two studies using retailored interventions to have the largest effect sizes

(.32 and .26), but moderator analysis could not be pursued due to sample size. Fail safe N was 9 and Egger's regression intercept was nonsignificant ($B = 1.15$ (2.44), $t = .47$, $df = 3$, $p = .33$), and trim and fill imputed no values, suggesting some confidence in the findings.

Vegetable Intake – Assessment or Minimal Intervention versus Tailored and Retailored

For the four vegetable intake outcomes the mean effect size was $g = .07$ (.05), $p = .17$. Results ranged from -.20 to .31 with significant heterogeneity present where $\nu = .04$, $Q(3) = 12.69$, $p = .005$ (see Table 11 and Table 13). Despite the small sample size, moderator analysis was pursued in this case. A significant trend was found for number of interventions where two or more interventions significantly increased the effect size ($B = .33$ (.11), $z = 3.04$, $p = .0002$), leaving a nonsignificant residual ($Q(2) = 3.47$, $p = .18$). Interventions using more than one contact had the largest effect sizes (.31 and .11). Given no significant difference, fail safe N was 0, yet Egger's regression was not significant ($I = -17.7$ (11.96), $t = 1.48$, $df = 2$, $p = .14$), but this may be an artifact of the sample size. Examination of the funnel plot suggested no bias for publication of small sample size studies with large effects and trim and fill imputed no values. See Table 11.

Comparison of Fruit versus Vegetable Intake

For all for studies measuring fruit and vegetable intake separately, a trend for larger effects was found for fruit intake, suggesting fruit intake is easier to increase. This difference was not significant ($Q_b(1) = 1.57$, $p = .21$), but given that four of five outcomes are dependent, the standard errors may be erroneously large.

Fiber Intake

Four studies intervened upon increased amounts of dietary fiber. Effect sizes for these interventions ranged from .18 to .93. Given the small sample size of the study showing an effect of .93, it had little weight and removing it only reduced the overall effect to .29 so it was retained. Therefore the mean effect size was $g = .34$ (.09), $p = .0001$. Studies were homogenous where $\nu = .03$, $Q(3) = 5.40$, $p = .15$ (see Table 11 and Table 13). Moderators were not examined given the small number of studies. Fail safe N was 16 and 3 with Rosenthal's and Orwin's methods respectively, but Egger's regression ($I = 4.55$ (.86), $t = 5.27$, $df = 2$, $p = .02$) was significant suggesting publication bias. Such bias would be expected given the publication of one small N study with a large effect. Imputation of one study reduced the overall mean effect to $g = .29$, the same result as if the study had been removed. See Table 11.

Table 11: Fiber, Fruit and Vegetable Results

Comparison	k	Model	g	SE	95% CI	Z	p	Q	P (df)	Fail Safe N R/O
Fruit /Vegetable Assess or minimal v. re/tailored	9	Fixed	.18	.02	.13-.23	7.37	.0001	16.48	.04(8)	66/8
		Random	.20	.03	.15-.25	7.23	.0001			
	6	Fixed	.11	.04	.03-.19	2.62	.009	11.65	.04(5)	
		Random	.11	.07	-.02-.23	1.59	.11			
	3	Fixed	.22	.03	.16-.28	7.23	.001	.01	.99(2)	
		Random	.22	.03	.16-.28	7.23	.001			
Fruit	5	Fixed	.17	.05	.07-.27	3.24	.001	3.92	.42(4)	9/0
		Random	.17	.05	.07-.27	3.24	.001			
Vegetables	4	Fixed	.07	.05	-.03-.22	1.37	.18	12.69	.005(3)	0/0
		Random	.06	.11	-.16-.28	.52	.60			
Fiber	4	Fixed	.34	.09	.17-.52	3.93	.0001	5.40	.15 (3)	16/3
		Random	.38	.12	.14-.62	3.11	.002			

Table 12: Minimal Intervention v. Tailored and Retailored Intervention for Fruit and Vegetable Intake

Study name	Comparison	Outcome	Time point	g	SE	Z	P
Anderson et al 2001	Assess v Tailored	Fruit & Veg/day	6	0.20	0.16	1.28	0.20
Campbell et al 2002	Assess v Tailored	Fruit & Veg/day	18	0.22	0.09	2.58	0.01
Prochaska et al 2005	Assess v Tailored+Manl	Fruit & Veg/day	12,24	0.22	0.03	6.63	0.00
Retailored Fixed				0.22	0.03	7.23	0.00
Retailored Random				0.22	0.03	7.23	0.00
Campbell & Burnhardt 1999	Assess v Tailored	Fruit & Veg/day	12	0.30	0.12	2.58	0.01
Campbell et al 1994	Assess v Tailored	Fruit & Veg/day	4	0.00	0.12	0.00	1.00
Campbell et al 2004	Assess v Tailored	Fruit & Veg/day	2	0.17	0.11	1.51	0.13
Heimendinger et al 2005	Brochure v Tailored	Fruit & Veg/day	12	0.12	0.07	1.81	0.07
Kreuter & Sugg 2005	Assess v Tailored	Fruit & Veg/day	6,18	-0.19	0.11	-1.67	0.10
Lutz et al 1999	Assess v Tailored	Fruit & Veg/day	6	0.22	0.12	1.88	0.06
Tailored Fixed				0.11	0.04	2.62	0.01
Tailored Random				0.10	0.07	1.60	0.11
Overall Fixed				0.18	0.02	7.37	0.00
Overall Random				0.20	0.03	7.23	0.00

Table 13: Minimal Intervention v. Re/Tailored on Fruit, Vegetable, and Fiber Intake

Study name	Comparison	Outcome	Time point	g	SE	Z	P
Brug et al 1996	Brochure v Tailored	Fruit /day	1	0.01	0.11	0.05	0.96
Brug et al 1998	Brochure v Retailored	Fruit /day	2	0.26	0.10	2.72	0.01
Brug et al 1999	Bhvr Feedback v Tailored	Fruit /day	1	0.14	0.11	1.28	0.20
Delichatsios et al 2001	Exercise v Retailored	Fruit /day	3	0.32	0.19	1.65	0.10
Oenema et al 2005	Assess v Tailored	Fruit /day	1	0.19	0.12	1.62	0.11
Fixed				0.17	0.05	3.24	0.00
Random				0.17	0.05	3.24	0.00
Brug et al 1996	Brochure v Tailored	Veg /day	1	0.00	0.11	0.00	1.00
Brug et al 1998	Brochure v Retailored	Veg /day	2	0.31	0.10	3.22	0.00
Brug et al 1999	Bhvr Feedback v Tailored	Veg /day	1	-0.20	0.11	1.81	0.07
Oenema et al 2005	Assess v Tailored	Veg /day	1	0.11	0.12	0.95	0.34
Fixed				0.07	0.05	1.37	0.17
Random				0.06	0.11	0.53	0.60
Anderson et al 2001	Assess v Retailored	Fiber g/day	6	0.37	0.16	2.33	0.02
Burnett et al 1989	Brochure v Tailored	Fiber g/day	3	0.93	0.30	3.15	0.00
Delichatsios et al 2001	Exercise v Retailored	Fiber g/day	6	0.37	0.19	1.96	0.05
Elder et al 2005	Targeted Brochure v Tailored	Fiber g/day	3	0.18	0.14	1.28	0.20
Fixed				0.34	0.09	3.93	0.00
Random				0.38	0.12	3.11	0.00

Exercise/Physical Activity

Outcomes

Studies reported exercise outcomes using five different outcome measures: Percent reaching Action or Maintenance stages, percent making any stage progress, percent meeting CDC exercise criteria, percent making any increase in exercise, and amount of activity measured by seven-day physical activity recall. Two studies reported outcomes on three of these measures concurrently. Bock et al. and Pinto et al. (2001; 2002) reported results in terms of percent reaching Action or Maintenance stages, percent meeting CDC exercise criteria, and seven day activity recall. There were no significant differences among these outcomes, although there was a trend for the seven-day recall to show smaller effects. Thus, when multiple outcomes are reported, CDC criteria will be preferred for the analysis, followed by percent reaching Action or Maintenance, and finally, seven-day activity recall.

Table 14: Comparison of Physical Activity Outcome Measures

Outcome	k	Hedges g	SE	Z	p
% A or M	6	.27	.07	3.76	.0001
% CDC	5	.31	.08	3.90	.0001
7 Day PAR	4	.23	.07	3.22	.0001

Results by Comparison

Physical Activity – Percent Reaching Criteria

Eleven studies measured outcomes in terms of percentage reaching CDC criteria, percent reaching Action/Maintenance stages, or mean physical activity recall. Six studies employed a retailored intervention, three employed a tailored or retailored

intervention plus counselor contact, and two employed a one-time tailored intervention. Due to small sample sizes, all interventions were combined for calculation of an overall effect size. Four studies measured outcomes at more than one timepoint, and the mean of these effects was used as the overall measure for each. After combining timepoints, effect sizes ranged from $g = .06$ to $.49$. One outlier was present where $g = .72$ (Bock et al., 2001, 6-mo) which appeared due to an intense intervention and was kept, as it entered into calculating the mean for that study with two additional timepoints. Overall the mean effect size for this group of studies under the fixed effect model was $g = .20 (.04)$, $p = .0001$ and $g = .24 (.05)$, $p = .0001$ using the random effects model (see Table 14 and Table 15).

The Q test for homogeneity of variance was not significant where $\nu = .007$, $Q(10) = 14.32$, $p = .16$. In terms of moderators, five studies employed proactive recruitment where $g = .22(.06)$ and six, reactive where $g = .19(.05)$, but studies did not differ ($Q_b = .11$, $p = .74$). Studies were conducted at seven different sites, preventing meaningful comparison. Retention rate and mean age were not related to effect size, but percent female was moderately related using $p < .10$ as the significance level ($B = -.39 (.23)$, $Z = -1.69$, $p = .09$), which reduced the residual to nonsignificance ($Q = 11.36$, $p = .25$). Percent minority and recruitment rate could not be assessed due to missing data.

Figure 6: Regression of Percent Female of Hedges's g , Physical Activity Criteria

Since studies varied by outcome timepoint and four provided multiple outcomes, outcomes were examined by timepoint grouping. Outcomes were grouped into three categories: 1-3, 4-6, and 7+ months. Five outcomes comprised each category. Due to dependence, statistical significance could not be examined across timepoint categories. This comparison contains a small number of studies, creating large confidence intervals that overlap, however, suggesting no significant differences over time. Examination of means suggest that strongest effects were found from 1-3 months ($g = .38$), followed by 4-6 months ($g = .31$), then 7+ months ($g = .12$). See Table 17 and Figure 7. Fail safe N was calculated to be 100 and 12 with Rosenthal's and Orwin's methods respectively. Both Egger's regression ($I = 1.91 (.58)$, $t = 3.29$, $df = 9$, $p = .005$) and trim and fill suggested an effect of publication bias. Trim and fill imputed six studies to the left of the mean reducing the overall effect to $g = .14$.

Figure 7: Mean ES by Timepoint for Physical Activity Criteria

Table 15: Physical Activity Mean ES

Comparison	k	Model	g	SE	95% CI	Z	p	Q	p (df)	Fail Safe N R/O
% Criteria										
Assess or minimal v. re/tailored	11	Fixed	.20	.04	.13-.27	5.55	.001	14.32	.16(10)	100/1 2
		Random	.24	.05	.14-.33	4.88	.001			
% Stg Progress										
Assess or minimal v. re/tailored	9	Fixed	.14	.05	.05-.23	2.99	.001	15.71	.05(8)	26/5
		Random	.18	.07	.04-.32	2.5	.01			

Table 16: Minimal Intervention v. Re/Tailored on Reaching Physical Activity Criteria

Study name	Comparison	Outcome	Time point	g	SE	Z	P
Bock et al 2001	Brochure v Retailored+Manl	% CDC Exercise	1,3,6,12	0.49	0.19	2.58	0.01
Campbell et al 2002	Assess v Retailored	7 Day PAR	18	0.06	0.09	0.74	0.46
Greaney et al 2007	Assess v Retailored+Manl+Call	% A or M Ex	12,24	0.08	0.06	1.18	0.24
Jacobson et al 2004	Advice v Multiple Tailored + Call	% A or M Ex	12	0.28	0.15	1.82	0.07
King et al 2006	Bhvr Feedback v Tailored+MI	7 Day PAR	2	0.38	0.12	3.29	0.00
Kosma et al 2005	Assess v Tailored	Long-term PAR	1	0.38	0.24	1.60	0.11
Krcuter & Strecher 1996	Assess v Tailored	% CDC Exercise	6	0.35	0.24	1.45	0.15
Napolitano et al 2003	Assess v Retailored	7 Day PAR	3	0.36	0.28	1.27	0.20
Pinto et al 2002	Assess v Retailored	% CDC Exercise	3,6	0.25	0.16	1.57	0.12
S.Johnson et al 2006	Assess v Retailored+Manl	% A or M Ex	6,12,18	0.40	0.20	2.02	0.04
Vandelanotte et al 2005	Assess v Retailored+Manl	% CDC Exercise	6	0.27	0.09	3.09	0.00
Fixed				0.20	0.04	5.55	0.00
Random				0.24	0.05	4.88	0.00

Table 17: Minimal Intervention v. Re/Tailored on Reaching Physical Activity Criteria by Timepoint Category

Study name	Comparison	Outcome	Time point	g	SE	Z	p
Bock et al 2001	Brochure v Retailored+Manl	CDC Exercise	3	0.33	0.16	1.99	0.05
King et al 2006	Bhvr Feedback v Tailored+MI	7 Day PAR	2	0.38	0.12	3.29	0.00
Kosma et al 2005	Assess v Tailored	Long-term PAR	1	0.38	0.24	1.60	0.11
Napolitano et al 2003	Assess v Retailored	7 Day PAR	3	0.36	0.28	1.27	0.20
Pinto et al 2002	Assess v Retailored	CDC Exercise	3	0.44	0.16	2.78	0.01
Fixed				0.38	0.07	5.15	0.00
Random				0.38	0.07	5.15	0.00
Bock et al 2001	Brochure v Retailored+Manl	CDC Exercise	6	0.72	0.21	3.44	0.00
Kreuter & Strecher 1996	Assess v Tailored	CDC Exercise	6	0.35	0.24	1.45	0.15
Pinto et al 2002	Assess v Retailored	CDC Exercise	6	0.06	0.16	0.36	0.71
S.Johnson et al 2006	Assess v Retailored+Manl	% A or M Ex	6	0.47	0.19	2.41	0.02
Vandelanotte et al 2005	Assess v Retailored+Manl	CDC Exercise	6	0.27	0.09	3.09	0.00
Fixed				0.31	0.06	4.71	0.00
Random				0.34	0.10	3.37	0.00
Bock et al 2001	Brochure v Retailored+Manl	CDC Exercise	12	0.43	0.19	2.18	0.03
Campbell et al 2002	Assess v Retailored	7 Day PAR	18	0.06	0.09	0.74	0.46
Greaney et al 2007	Assess v Retailed+Manl+Call	% A or M Ex	12	0.06	0.06	0.91	0.36
Jacobs et al 2004	Advice v Multiple Tailored + Call	% A or M Ex	12	0.28	0.15	1.82	0.07
S.Johnson et al 2006	Assess v Retailored+Manl	% A or M Ex	12	0.34	0.21	1.64	0.10
Fixed				0.11	0.05	2.49	0.01
Random				0.15	0.06	2.31	0.02

-2.000 -1.500 -1.000 -0.500 0.000 0.500 1.000 1.500 2.000

Physical Activity – Percent Making Progress

Nine studies measured outcomes in terms of percent of participants making stage progress or increasing amount of exercise. Stage progress refers to percent of participants moving at least one stage of change forward on the continuum from Precontemplation to Contemplation to Preparation to Action to Maintenance as specified by the Transtheoretical Model. Despite the dichotomous nature of this outcome, the logit transformation was taken to enable comparison with other physical activity outcomes employing Hedges g . Four studies measured outcomes on more than one occasion. Overall, however, seven outcomes were available for 1-3 month outcome, two for 4-6 month outcome and 3 for 12 month outcome and thus comparisons by timepoint were not feasible. The mean of multiple timepoints was included in the combined effect size analysis. Effect sizes ranged from .05 to .66. The mean effect under the fixed effects model was $g = .14$ (.05), $p = .001$ and $g = .18$ (.07), $p = .01$ under the random effects model. Significant heterogeneity was present among studies where $\nu = .02$, $Q(8) = 15.54$, $p = .05$ (see Table 15 and Table 18). Examination of studies revealed that studies with the two highest effect sizes both employed an intensive, interactive web-based intervention (Kosma, Cardinal, & McCubbin, 2005; Napolitano et al., 2003). These studies also employed reactive recruitment strategies. When grouped by mode of intervention, within group heterogeneity was nonsignificant ($Q_w(6) = 6.96$, $p = .32$), but between groups differences were found ($Q_b(2) = 8.74$, $p = .01$). A significant regression was found for percent female ($B = -.54$ (.27), $Z = -2.03$, $p = .04$) reducing the residual to nonsignificance ($Q = 11.71$, $p = .11$) (see Figure 8). Rosenthal's fail safe N was 26 and Orwin's 5, suggesting the

tenuous nature of this effect. Trim and fill imputed two values, reducing the effect size to $g = .12$ and Egger's regression was not significant, probably due to large standard error ($I = 1.55(1.04)$, $t = 1.50$, $df = 7$, $p = .09$).

Figure 8: Regression of Percent Female on Hedges g, Physical Activity Progress

Table 18: Minimal Intervention v. Re/Tailored on Increasing Physical Activity

Study name	Comparison	Outcome	Time point	g	SE	Z	p
Bull & Jamrozik et al 1999	Brochure v Tailored	% Increasing Ex	1,6,12	0.02	0.11	0.21	0.83
Bull & Kreuter et al 1999	Brochure v Tailored Blvr Feedback v Tailored	% Increasing Ex	3	0.10	0.28	0.38	0.71
Cardinal et al 1995	% Stg Progress	% Stg Progress	1,7,12,24	0.04	0.29	0.15	0.88
Greaney et al 2007	Assess v Retailled+Manl+Call Advice v Multiple Tailored + Call	% Stg Progress	12	0.11	0.08	1.29	0.120
Jacos et al 2004	% Increasing Ex	% Increasing Ex	12	-0.05	0.10	-0.50	0.62
Kosma et al 2005	Assess v Tailored	% Stg Progress	1	0.57	0.24	2.34	0.02
Marcus et al 1998	Brochure v Multiple Tailored	% Stg Progress	3	0.25	0.14	1.78	0.04
Napolitano et al 2003	Assess v Retailored	% Stg Progress	1,3	0.66	0.30	2.21	0.03
Peterson et al 1999	Brochure v Tailored	% Stg Progress	3	0.42	0.14	3.00	0.00
Fixed				0.14	0.05	2.99	0.00
Random				0.18	0.07	2.50	0.01

Smoking

Choice of outcome measure is an important and somewhat controversial decision in regard to smoking cessation studies. FDA guidelines recommend 28-day abstinence or longer as the preferred method of assessing efficacy (FDA, 1995). Many studies, however, do not employ this method, using either 24-hour or 7-day abstinence as outcome measures. Studies in the meta-analysis sample include outcomes assessing smoking using complete abstinence at 24-hour, 7-day, 28-day, 10-week, and 6-month increments. If such outcomes could be combined this would result in a larger sample for analysis. In an analysis of smoking outcome measures, Velicer et al. (2004) found that 24-hour, 7-day point prevalence, and 30-day prolonged abstinence measures showed correlations of at least .98 with each other among a series of three similar studies. The six-month continual abstinence measure showed a correlation of .82 with these measures. They note two main problems with using prolonged abstinence as an outcome measure: (1) that it ignores the usual pattern of quitting in which people relapse multiple times and (2) that it does not account for delayed quitting in the sample. Thus prolonged abstinence rates tend to decrease across time, whereas the other measures show increases. This pattern will be examined in this analysis to determine if it replicates across disparate studies. In the present analysis 24-hour, 7-day point prevalence, and 28-day prolonged abstinence outcomes will be analyzed as if they are equivalent. Where studies report more than one of these measures for each timepoint, 24-hour point prevalence will be preferred, followed by 7-day point prevalence, and then by 30-day prolonged abstinence. Six month and 10-week

sustained abstinence will be examined separately. Since all outcomes are reported in dichotomous format, log odds will be reported as the effect size measure.

Smoking Outcomes: 6 Month Abstinence

For smoking behavior, 10 studies reported outcomes for prolonged abstinence (10 weeks, 6-months or Action or Maintenance outcomes). For each study the mean effect for multiple timepoints was used since few differences were found across outcome assessment timepoints ($Q_b = .69, p = .71$). Effect sizes ranged from .12 to .75. For the fixed effects and random effects models the overall effect was $LO = .45 (.06), p = .001$ and $LO = .49 (.09), p = .001$ (see Table 19 and Table 20). Studies were homogenous where $\nu = .01, Q(9) = 14.47, p = .11$. Nevertheless differences were found between tailored ($LO = .34$) and retailored studies ($LO = .65$) where $Q_b(1) = 5.16, p = .02$. Recruitment rate was also a significant predictor where $B = 3.21(.94), p = .0001$ (see Figure 9) and retention rate $B = -2.03(.78), p = .01$ (see Figure 10), but not mean age $B = -.01(.01), p = .42$. No differences were found between proactive ($LO = .52$) or reactive ($LO = .41$) recruitment strategy $Q_b = .60, p = .44$ or by study group $Q_b = 2.84, p = .42$. Rosenthal's fail safe N was 117 and Orwin's 40 and Egger's regression was not significant ($I = 11.16 (.86), t = 1.35, df = 8, p = .22$). Trim and fill indicated slight publication bias, suggesting imputation of three studies to the left of the mean, reducing overall effect size slightly to $LO = .40$ (95% CI = .28 - .52) with the fixed effects model.

Figure 9: Regression of Recruitment Rate on Mean ES 6-mo Abstinence

Figure 10: Regression of Retention Rate on Mean ES 6-mo Abstinence

Table 19: Smoking Mean ES 6-Mo Abstinence

Comparison	k	Model	LO	SE	95% CI	Z	p	Q	P (df)	Fail Safe N R/O
Assess or minimal v. re/tailored	10	Fixed	.45	.06	.32-.57	6.97	.001	14.49	.11(9)	117/40
		Random	.49	.09	.31-.67	5.28	.001			
Tailored	4	Fixed	.34	.18	.19-.50	4.38	.001	7.33	.06	
		Random	.32	.14	.04-.60	2.25	.025			
Retailored	6	Fixed	.65	.11	.43-.87	5.88	.001	2.00	.85	
		Random	.65	.11	.43-.87	5.88	.001			

Table 20: Minimal Intervention v. Re/Tailored on 6-Month Smoking Abstinence

Study name	Comparison	Outcome	Time pnt	LO	SE	Z	p
Assess v							
Prochaska et al 2004	Retailored+Manl	6 mo Quit	12,24	0.67	0.35	1.90	0.06
Borland 2004	Brochure v Retailored	6 mo Quit	12	0.66	0.23	2.91	0.00
Curry et al 1991	Brochure v Retailored	6 mo Quit	12	0.75	0.32	2.30	0.02
Curry et al 1995	Man v Tailored+Man	6 mo Quit	12,21	0.21	0.46	0.46	0.65
Prochaska et al 2001	Assess v Retailored	6 mo Quit	12,18,24	0.53	0.18	2.91	0.00
Prochaska et al 2001a	Assess v Retailored+Manl	6 mo Quit	12,18,24	0.62	0.35	1.78	0.08
Shiffman et al 2000	Man+Patch v Tailored+Patch	10 wk Quit	3	0.57	0.12	4.90	0.00
Shiffman et al 2001	Man+Patch v Tailored+Patch	10 wk Quit	12	0.12	0.12	1.00	0.32
Velicer et al 1999	Man v Retailored	6 mo Quit	12,18	1.16	0.42	2.75	0.01
Velicer et al 2006	Man v Man+Tailored+Patch	6 mo Quit	10,20,30	0.27	0.24	1.14	0.26
Fixed				0.45	0.06	6.97	0.00
Random				0.49	0.09	5.28	0.00

Smoking Outcomes: 24 hour, 7-Day and 28-Day Quit

Overall Analysis

For the overall analysis two studies (Streicher et al., 2005; Velicer et al., 1999) employed both a tailored and a retailored comparison group. The retailored was chosen for the overall analysis as more studies overall employed this method. Two small-N studies showed large effect sizes ($LO = 1.16$) and could be considered outliers. The first, (Dijkstra, 2005) employed interactive computer and (Lipkus, Lyna, & Rimer, 1999) employed a physician advice component whereas other studies did not. These were not considered representative and were removed from analysis. Effect sizes ranged from $LO = -0.41$ to 1.10 . The mean effect for the 23 studies included was $LO = .28 (.03)$, $p = .001$ with fixed effects and $LO = .28 (.04)$ with random effects modeling. Heterogeneity among studies was not found ($\nu = .001$, $Q(22) = 24.83$, $p = .31$) (see Table 21 and Table 22). No differences between studies was found for proactive versus reactive recruitment strategy $Q_b = .13 (1)$, $p = .71$, tailored or retailored modality $Q_b = .07(1)$, $p = .79$, or study group $Q_b = 8.2(10)$, $p = .60$. Among continuous predictors no differences were found for gender $B = -.04 (.14)$, $p = .76$, mean age $B = .004 (.004)$, $p = .39$, or for retention rate $B = .10 (.16)$, $p = .54$. Minority status and recruitment rate were not examined due to greater than 10% missing data. Publication bias was minimal for these studies, showing symmetrical funnel plot and with trim and fill suggesting no imputed studies. Eggers regression was not significant ($I = .05(.46)$, $df = 21$, $p = .91$) and Rosenthal's fail safe $N = 297$ and Orwin's = 40 using $LO = .09$ as the minimal effect for clinical nonsignificance.

Re/Tailored versus Re/Tailored + Counselor Calls

Six studies included tailored or retailored intervention alone and in combination with counselor calls, allowing for comparison of the additive effect of calls over print tailoring. The mean effect with the fixed and random effects models was $LO = .20(.09)$ and $.19 (.13)$ respectively. Due to large standard error, the random effects model is nonsignificant. Heterogeneity was not present were $\nu = .04$, $Q(5) = 8.22$, $p = .14$ (see Table 21 and Table 23). Compared to retailoring alone, the effect of adding counselor calls resulted in increased effects at short-term follow up, but smaller effects over time (see Figure 12).

Table 21: Smoking Mean ES: 24-Hr, 7-Day, and 28-Day Quit

Comparison	k	Model	LO	SE	95% CI	Z	p	Q	p (df)	Fail Safe N R/O
Assess or minimal v. re/tailored	23	Fixed	.28	.03	.21-.34	8.33	.001	24.83	.31(22)	297/40
		Random	.28	.04	.21-.35	7.39	.001			
Tailored	10	Fixed	.26	.04	.18-.35	6.02	.001	13.33	.15(9)	
		Random	.26	.07	.13-.39	3.97	.001			
Retailored	13	Fixed	.28	.05	.19-.38	5.77	.001	12.87	.54(12)	
		Random	.28	.05	.19-.38	5.77	.001			
Proactive	13	Fixed	.29	.05	.19-.40	5.4	.001	14.57	.27(12)	
		Random	.29	.06	.17-.42	4.71	.001			
Reactive	10	Fixed	.27	.04	.18-.35	6.35	.001	10.12	.34(9)	
		Random	.27	.05	.17-.36	5.58	.001			
Re/Tailored v Re/Tailored + Calls	6	Fixed	.20	.09	.01-.38	2.12	.03	8.22	.14(5)	
		Random	.19	.13	-.06-.44	1.52	.13			

Table 22: Minimal Intervention v. Re/Tailored on Short Term Smoking Abstinence

Study name	Comparison	Outcome	Time pnt	LO	SE	Z	p
Aveyard et al 2003	Brochure v Retailored+Manl	7 Day Quit	12	0.26	0.18	1.40	0.16
Borland 2004pp	Brochure v Retailored	7 Day Quit	12	0.06	0.18	0.34	0.74
Borland et al 2003	Man v Retailored	7 Day Quit	3,6,12	0.10	0.18	0.55	0.58
Campbell et al 2002	Assess v Retailored	7 Day Quit	18	0.00	0.16	0.00	1.00
Curry et al 1991	Brochure v Retailored	24 Hour Quit	3	0.32	0.16	1.98	0.05
Curry et al 1995	Man v Tailored+Man	7 Day Quit	3,12,21	-0.02	0.31	-0.07	0.94
Dijkstra et al 1999	Man v Tailored	7 Day Quit	6	0.24	0.53	0.45	0.65
Etter et al 2004	Assess v Retailored	7 Day Quit	6,24	0.46	0.15	3.16	0.00
Etter et al 2005	Brochure v Tailored	7 Day Quit	3	0.22	0.07	3.12	0.00
Kreuter & Strecher 1996	Assess v Tailored	7 Day Quit	6	-0.41	0.53	-0.77	0.44
Lennox et al 2001	Assess v Tailored	28 Day Quit	6	0.31	0.28	1.08	0.28
Owen et al 1989	Brochure v Tailored	7 Day Quit	1,3,9	-0.19	0.42	-0.45	0.65
Prochaska et al 2001	Assess v Retailored	24 Hour Quit	6,12,18,24	0.31	0.11	2.75	0.01
Prochaska et al 1993	Brochure v Retailored+Manl	24 Hour Quit	18	0.57	0.34	1.69	0.09
Prochaska et al 2001a	Assess v Retailored+Manl	24 Hour Quit	6,12,18	0.39	0.23	1.68	0.09
Prochaska et al 2004	Assess v Retailored+Manl	24 Hour Quit	12,24	0.37	0.27	1.39	0.16
Prochaska et al 2005	Assess v Retailored+Manl	24 Hour Quit	12,24	0.46	0.19	2.45	0.01
Shiffman et al 2000	Man+Patch v Tailored+Patch	28 Day Quit	1	0.55	0.12	4.51	0.00
Strecher & Shiffman 2005	Man+Patch v Tailored+Patch	28 Day Quit	3	0.26	0.08	3.42	0.00
Strecher et al 1994	Assess v Tailored	7 Day Quit	6	1.10	0.47	2.35	0.02
Strecher et al 2005	Brochure v Retailored	7 Day Quit	12	0.26	0.25	1.05	0.30
Valicic et al 1999	Man v Retailored	24 Hour Quit	6,12,18	0.59	0.24	2.42	0.02
Valicic et al 2006	Man v Man+ Tailored+Patch	24 Hour Quit	10,2030	0.10	0.17	0.56	0.58
Fixed				0.28	0.03	8.33	0.00
Random				0.28	0.04	7.39	0.00

Table 23: Re/Tailored versus Re/Tailored + Counselor Calls Smoking Cessation

Study name	Comparison	Outcome	Time pnt	LO	SE	Z	p
Borland et al 2003	Retailored v Retailored + Call	7 Day Quit	3,6,12	0.36	0.17	2.15	0.03
Curry et al 1995	Tailored+Manl v Tailored+Manl+Call	7 Day Quit	3,12,21	0.59	0.34	1.76	0.08
Lipkus et al 1999	Tailored+Advice v Tailored+Advice+Call	28 Day Quit	16	-0.71	0.45	-1.57	0.12
Prochaska et al 1993	Retailored v Retailored+Manl+Call	24 Hour Quit	18	0.43	0.29	1.47	0.14
Prochaska et al 2001a	Retailored v Retailored+Manl+Call	24 Hour Quit	6,12,18	0.13	0.22	0.59	0.55
Velicer et al 2006	Man+tailored+patch v Tailored+Man+patch+Call	24 Hour Quit	10,20,30	0.02	0.17	0.10	0.92
Fixed				0.20	0.09	2.12	0.03
Random				0.19	0.13	1.52	0.13

Analysis by Timepoint

Since most studies assessed outcomes at more than one timepoint, studies were examined across timepoints. Timepoints cannot be compared with formal statistical tests due to dependence, in which case the null is rarely rejected, but can be presented for determination of trends of effects over time. Examination of overlapping confidence intervals allows estimation of significance, however, and suggests that the trend is not significant with the sample size included here. The mean effect at 1-3 month assessment was $LO = .27 (.04)$, at 4-6 months $LO = .35 (.08)$, at 7-12 months $LO = .31 (.05)$, at 13-23 months $LO = (.24)$ and at 24 month assessment and longer $LO = .17 (.06)$. Figure 11 shows the effect size trend by timepoint with 95% confidence intervals.

Figure 11: Mean ES for Smoking Cessation Interventions by Timepoint

**Figure 12: Long Term Effect for Smoking Cessation for Retailored Print versus
Retailored Print + Counselor Calls**

CHAPTER 6: DISCUSSION

Mammography

The overall mean effect size for tailored or retailed mammography intervention suggests about a 24% increase in effectiveness for these interventions compared to minimal contact. These studies showed significant heterogeneity, however, which was explained by positive correlations among effect size, retention rate, and minority status. Studies showing greater retention rates may result in greater success since high retention facilitates more participants receiving the intervention. Increased success for interventions aimed at minority populations also makes theoretical sense given the base rates of minorities getting regular mammography are lower than non-minorities. The interventions would have more relative success moving a group from 55% mammography status, to 70% status than moving a higher SES from 65% to 70% getting mammograms. The effect of tailored versus retailed interventions could not be assessed for mammography given that only two studies employed a retailed component. All studies recruited participants proactively and no differences were found across theoretical orientation or study group. The effects of the interventions declined across time for both the tailored and tailored plus counselor call condition, with the largest drop from six to 12 months post-baseline. Follow-up past 12 months was not available for the counselor call condition to compare to the tailored intervention alone, which did maintain effects of about $OR = 1.22$ in the three studies assessing at periods greater than 13 months. Thus, it appears that interventions aimed at increasing mammography use maintain at least some effect over time.

Adding counselor calls to the tailored interventions appears to result in 2.8 times larger effects for this behavior. This difference could arise from two sources: 1). studies employing calls being of better quality and/or 2). Effect of retailoring that calls provided. It does appear that the interventions that employed additional counselor call interventions were of better quality. When the tailored intervention components of the studies that employed calls were compared alone, the LO = .37 (OR = 1.45), which is greater than the overall mean for all tailored interventions LO = .24, suggesting that the tailored component of the call studies was of greater effectiveness alone. Additionally, since only two studies employed a retailored component, the effect of calls over retailoring could not be assessed. It is possible that counselor calls provided an updated assessment and feedback component that would be independent of intervention modality. It does appear, however, that adding calls to tailored intervention does increase effectiveness for mammography behavior by 1.7 times.

The relative risk calculated for mammography screening interventions is 1.12. Examination of the relative risk facilitates interpretation of the practical effects of intervention effects. Relative risk employs the overall sample size per group as the denominator and thus does not compare treatment versus intervention groups. Relative risk can be used to show the amount of change over the entire population if an intervention were provided and thus usually will be lower than the odds ratio. Subtracting the relative risk from 1 yields relative risk reduction (RRR) in percentage terms, which would be 12% in this instance. If the intervention were provided, therefore, one would expect a 12% change in mammography behavior overall. Since effect size relies on difference scores, it is necessary to base interpretation on the

actual base rates of the behaviors under question. The absolute rate of women getting mammography ranges from 63% to 75% depending on SES. These interventions offer the possibility of increasing mammography rates of people not already getting mammography by 7-10% in the population (see Table 24). Given that the relative risk of breast cancer mortality decreases about 23% for women over 50 who get regular mammograms (Agency for Healthcare Research and Quality, 2002), increasing rates by 10% could decrease the 41,000 mortality rate by about 2.3% or 943 persons.

Table 24: Risk Reduction and Number Need to Treat for Mammography

Screening

	Absolute Event rate	RRR	With Tx	Absolute RR	NNT
Tailored	70%	12%	80.4%	10.4	9
Tailored + Call	70%	45%	100%	31.5	3
Call alone	70%	14%	84%	9.8	10
Call over Tailored	70%	15%	80.5%	10.5	10

Diet

The mean ES for all tailored or retailored dietary fat reduction studies was $g = .22$, a fairly large population effect. Retailored interventions offered 60% greater effectiveness than tailored interventions. Effects were maintained over follow-up at 24 months averaged across all studies, but this should be interpreted with caution given that only two studies reported 24 month outcomes. Only two studies added counselor calls to re/tailored interventions, but four did include counselor call only conditions. Comparison of minimal intervention and counselor calls revealed effects similar to those of retailored print interventions. The effect for counselor calls, however, decreased across time versus the greater stability found for the re/tailored

modalities. Gender was shown to predict effects for fat reduction, suggesting a trend for studies to have smaller effect with increasing percentages of women included. Individual studies have found larger effects for decreasing fat intake for men than women (Kristal et al., 2000). The CDC estimates that in 2000 men and women both consumed about 11% of calories from fat with rates comparable between gender for the past 30 years (CDC, 2004). At least one study in the present analysis, however, did find and report higher baseline consumption of fat for men than for women (Armitage et al., 2001) and other studies have identified women as attending to health message to a greater extent than men (Stevens et al., 2003).

For fruit and vegetable intake, the mean effect was equivalent to that for dietary fat reduction. Retailored interventions showed about twice the effectiveness of tailored interventions. Effects increased from baseline to 6-month outcomes and either decreased or remained consistent at 12-months post-intervention. The small sample size of only four outcomes at 12 months in the presence of a large negative outlier may skew these results. Effect of counselor calls could not be assessed since only two fruit and vegetable studies employed this modality. Similar to dietary fat reduction, studies including more women showed smaller effects, again possibly because women already are eating improved diets. Indeed, recent estimates from the Behavioral Risk Factor Surveillance System found that only 22% of men but 32% of women eat at least three vegetables per day (CDC, 2007).

Analysis of studies intervening separately on fruit or vegetables or measuring them separately reveals small effects for vegetable intake ($g = .06$) alone, whereas the mean effect for fruit ($g = .17$) is more similar to that of fruit and vegetable intake

assessed together ($g = .23$). This suggests that interventions aiming to increase both fruit and vegetable intake have greater effects on increasing fruit versus vegetable consumption. The four studies focusing on increasing dietary fiber showed larger effects than the other dietary endpoints ($g = .38$). This could possibly be explained by the fact that fiber was an additional outcome in studies of fruit and vegetable intake. Increasing fruit and vegetable intake, therefore, also increases intake of fiber and the outcomes are likely dependent. Additionally, adding fiber to one's diet may be easier than adding fruit and vegetables. Fruit and vegetables requires addition of items formerly not consumed, whereas increasing fiber involves slight changes to common items such as choosing whole wheat over white bread.

Most studies employed the food frequency questionnaire developed by Block. Translated to that metric, a g of .22, a fairly large effect for population interventions, translates to a drop in fat score of 7 points. In terms of relative risk this represents a 30% decrease in dietary fat if provided on a population basis. High dietary fat intake combined with low fruit and vegetable intake is a significant contributor to overall obesity, the rates of which continue to increase (US Department of Health and Human Services, 2001). Decreasing fat by 30% and fruits and vegetables by 30% would no doubt help in combating obesity. Large weight reduction is not necessary to find benefit - decreasing overall weight even by 5-15% leads to significant decreases in risk of cardiovascular disease (US Department of Health and Human Services, 2001). Since such a large number of people are not engaging in healthy eating behavior, any intervention can have significant population impact. For example, since 73% of people do not eat five or more fruits or vegetable per day, increasing this rate by 30% would

increase the overall rate of consumption to 50% and meet the Healthy People goals for 2010 (CDC, 2007).

Physical Activity

The mean effect size for percent reaching criteria was $g = .24$, a large effect for population-based interventions. This represents a 39% increase in effect over the control group and a 28% increase in exercise if employed on a population. Recruitment strategy did not influence this outcome, but effect size was found to increase with the inclusion of fewer females in the studies. This suggests that physical activity interventions are more effective for men than women, which may be accounted for by the lower rates of activity by men (21%) than by women (26%) overall (CDC, 2005). A trend for decreasing effect size over time was found for percent reaching criteria, with a large drop between six and twelve month outcomes. It appears that the effect of the intervention decreases over time, but the drop is nonsignificant and the overall effect remains clinically meaningful where $g = .15$. A small to medium-size mean effect was found using percent of participants making progress as the outcome.

The CDC estimates that from 1994 to 2004 the percent of people engaging in physical activity at recommended levels decreased from 29.8% to 23.7%, with the largest decreases and absolute rates among white older men (CDC, 2005). Well-controlled studies estimate that engaging in regular physical activity reduces relative risk of death by about .70. The relative risk of mortality associated with inactivity is about 1.70 for men and .95 for women, which is fairly small compared to relative risks associated with smoking and heart disease, with risk equal to at least 3.0 (US

Department of Health and Human Services, 2001). The prevalence of physical inactivity, however, dramatically increases the population attributable risk. With 76% of the population not engaging in the protective behavior, even a small intervention impact garners consideration. Population interventions increasing the 76% inactivity rate by 28% would not only reduce the 10 year decline in activity but increase population rates to 45% engaging in regular physical activity, reducing relative risks for almost half of the population.

Smoking

Intervention effect measured by 24-hr, 7-day, and 28-day quit represented a 32% increase in smoking cessation in the treatment versus control group. This represents a 26% change in smoking if the interventions were applied to the population as a whole. For measurement of 6-month abstinence, the increase over the control group was 45%, representing a 46% effect at a population level. Thus, it may erroneously appear that measuring quit rates in terms of 6-month abstinence shows greater reductions in quitting. The large difference between measuring short and long-term outcomes most likely arises from the nature of proportional indices. Proportional measures are susceptible to misinterpretation due to confounding with overall sample size. For this group of studies, effect size decreases slightly over assessment timepoint measured continuously ($B = -.01, p = .09$) (see Figure 13). The proportions quit for treatment and control measured with short term quit rates are, for example, 25.6% and 14.4% quit versus 8.5% and 4.4% quit measured in terms of 6-month abstinence. The short term measures arise from proportions equaling 57/223 and 39/277 whereas the 6-month rates equal the result of 19/223 and 12/277. Examination of the data thus shows

that *absolute* rates of quitting are greater measured in terms of 24-hr quit, but not *relative* rates. This means that fewer people overall are going to have quit for 6 months continuously, but of those people, treatment has a larger effect over control. In other words, the treatment effect becomes greater over time, a common finding in tailored smoking cessation studies (Prochaska et al., 2001). Significant regression for retention rate on LO suggests that studies with higher retention show smaller effect sizes. This occurs since the results of smoking cessation take a longer time to show an effect. Lack of measuring at extended outcomes would thereby limit finding significant effects with tailored interventions.

Figure 13: Regression of Timepoint on Log Odds for Short Term Smoking Cessation

In terms of practical significance, about 21% of the US population was measured as smoking in 2005 (CDC, 2006). Reducing this by 25% would result in a 5% decrease in overall smoking rates, reducing the rate to about 16%. A 5% decrease in smoking would result in a 5% decrease in the \$75.5 billion spent on health care

costs annually or 3.8 billion dollars (CDC, 2002). A 5% decrease would also result in 9,300 fewer deaths from all cancer and 8,900 deaths from cardiovascular disease (CDC, 2002).

Summary of Predictions

Effectiveness of Tailored Interventions

Over all four behaviors the mean effect size was statistically significant for each comparison. Tailored interventions outperformed assessment-only or minimal interventions, with greater effects for retailored interventions. Table 25 presents a summary of effect sizes in odds ratio and relative risk format across all interventions and behaviors. Using Rossi's (2003) suggestions for estimating size of effects for population interventions, small, medium, and large effects in terms of odds ratios would be 1.32, 1.43, and 1.58. Table 26 summarizes results of the present study's main predictions across behaviors.

Table 25: Summary of Effect Sizes Across Behaviors and Comparisons

Comparison	Mammography		Diet Fat		Diet Fruit & Veg		Phys Activity		Smoking 6 mo quit		Smoking 24hr, 7d, 26d	
	OR	RR	OR	RR	OR	RR	OR	RR	OR	RR	OR	RR
Minimal v. Re/Tailored	1.23	1.08	1.45	1.31	1.44	1.30	1.39	1.28	1.58	1.45	1.32	1.26
Minimal v. Tailored	-	-	1.31	1.06	1.22	1.10	-	-	1.38	1.28	1.30	-
Minimal v. Retailored	-	-	1.56	1.39	1.50	1.32	-	-	1.91	1.71	1.32	-
Minimal v. Calls	1.45	1.14	1.50	1.35	-	-	-	-	-	-	-	-
Tailored v. Tailored + Calls	1.88	1.45	-	-	-	-	-	-	-	-	1.21	1.10

Table 26: Summary of Predictions and Findings by Behavior

Prediction	Mammography	Diet		Physical Activity % Criteria	Smoking Cessation	
		Fat	Fruit/Veg		Short term	6 mo abstinence
Tailored v. Retailored	2 tailored: Unable to compare	Retailored significantly greater	Retailored significantly greater	Unable to compare	Small effect for tailored	Retailored significantly greater
Effects over Time	Decreases over time, esp after 6mos	Increase to 6 mos, then consistent	Increase to 6 mos, then consistent	Decrease across time, esp after 6 mos	Increase to 6 mos then decrease across time	NA
Proactive/Reactive Recruitment Strategy	All proactive. Unable to compare.	1 reactive. Unable to compare.	All proactive. Unable to compare.	No differences	No differences	No differences
Effect of Calls	Added effect to tailored, sustained over time	Adding calls to tailored similar to tailored, decreased more than print tailored	2 used calls: Unable to compare	2 used calls: Unable to compare	Increased overall effect. Smaller effect than print shown over time.	Increased overall effect. Unable to compare across time
Differences by Theory	Too many categories. Unable to compare	Too many categories. Unable to compare	Too many categories. Unable to compare	Too many categories. Unable to compare	Too many categories. Unable to compare effect	Too many categories. Unable to compare
Differences by Study Group	Too many categories. Unable to compare	Too many categories. Unable to compare	Too many categories. Unable to compare	Too many categories. Unable to compare	Too many categories. Unable to compare	Too many variations. Unable to compare
Additional Moderators	Retention Rate (B = -1.41) Minority Status (B = .44)	For tailored % female (B = -.26)	For tailored % female (B = -.73)	% female (B = -.39)	None	Recruitment rate (B = 3.21) Retention rate (B = -2.03)

Differences among Modality of Intervention

Differences between Tailored or Retailored Interventions

The difference between tailored versus tailored interventions appears for smoking cessation, dietary fat reduction, and increasing fruit and vegetable intake such that retailoring almost doubled the effectiveness of tailored interventions. The greater effect could be explained by increased number of overall contacts that retailoring necessitates, but number of contacts regressed on effect size was not a significant predictor for any behavior. This suggests that periodical reassessment to enable updated feedback provides qualitatively meaningful improvement in interventions over and above increasing number of contacts. Sufficient data were available for smoking cessation to compare the effects of tailored versus tailored interventions across time. It appears that for shorter term outcomes, tailored interventions gave greater effects and that these endure more so than tailored interventions (see Figure 14). Results thus suggest that retailoring, despite resultant increased participant burden and effort involved, provides meaningful behavior change information that facilitates long-term maintenance.

Figure 14: Tailored v. Retailored Mean ES over Time, Smoking Cessation Short-Term Quit

Counselor Calls: Added Benefit?

The effect of calls varied across behavior and the limited number of studies that included this condition limited the comparisons possible. For mammography screening behavior, adding counselor calls to tailored intervention increases the odds of screening by 30%. It could be assumed that counselor calls functioned as retailored interventions, providing feedback updated to a person's current behavior change needs. Indeed, for dietary fat reduction, counselor calls showed a mean effect similar to that of the retailored interventions. The effect of retailored print intervention versus calls could not be assessed reliably as few mammography and diet studies employed sufficient comparison groups. Adding calls to tailored interventions increased the effectiveness for smoking cessation measured by 24-hr and 7-day quit by 16% initially, but resulted in smaller effects than tailored interventions alone over long-

term follow-up. For smoking cessation using 6-month abstinence, three studies compared calls alone to retailed interventions, showing an added benefit of calls of 14%. Using shorter abstinence outcomes, four studies showed a mean increase of 15% for calls alone over retailed interventions. A recent meta-analysis (Pan, 2006) of counselor calls alone for smoking cessation compared to no intervention found a mean effect of $OR = 1.46$, which is similar to that found in this study, suggesting the effectiveness of calls for smoking cessation, at least initially.

The effectiveness of counselor calls must be interpreted cautiously as it has been found that the effects of counselor call interventions are not maintained over time as are those of tailored interventions alone (Prochaska et al., 2001). Whenever possible this study compared the effects of retailed and tailored interventions versus calls over long-term follow-up. For mammography, counselor calls added to intervention effectiveness across all timepoints, although these interventions were not retailed. For dietary fat reduction, counselor calls functioned similarly to retailed interventions, with decreasing effectiveness over time. Previous predictions were maintained across disparate studies for smoking cessation such that calls added benefit initially to quit rates, but effects for calls declined more sharply than for tailored interventions alone. Thus it appears that counselor calls provide short-term efficacy, but less intensive print retailing fulfills the need for outcome maintenance, possibly due to reliance on the counselor rather than self-efficacy.

Effect Size Over Time

It was predicted that effect size would increase over time. In fact, across behaviors in which sufficient data were available, effect sizes decreased over time.

Across behaviors, the sharpest decreases in effect size were seen past 6-month follow-up. This was also accompanied by an increase in the error of measurement, as a surprising number of studies measured outcomes at relatively short outcome timepoints, often one to three months post-baseline. Such methodology limits the ability to detect long-term effects and makes little sense when attempting to measure behavioral change. Indeed such methodology assumes that the entire sample is ready to commence the behavior, an assumption that ignores stages of readiness. The hypothesis that tailoring methodology would improve since the first tailoring interventions were developed, resulting in larger effects was also tested by regressing publication date on effect size for each behavior. None of these regressions revealed a trend, significant or not.

Differences in Recruitment Strategy

It was predicted that proactive recruitment would result in smaller effect sizes, but reach larger numbers of people, resulting in greater impact. Due to small numbers of studies using reactive methodology for mammography and diet behaviors, this prediction could not be addressed. No differences in effect size by recruitment strategy were found for increasing physical activity or for smoking cessation.

Differences of Effect Size Among Behavior Change Theories

As predicted, theory was not a significant moderator in any comparison for any behavior. This could be due to the fact that many interventions employed either the TTM, did not refer to a particular theory, or combined components of various theories. Since most change theories incorporate similar variables into tailored feedback reports, finding differences in either number of variables intervened upon as a

continuous variable or among change theories in the small sample included here becomes difficult. In a related prediction, the multitude of study groups involved in creating tailored interventions increases the number of categories in some cases to the number of effects, thereby preventing statistical or even visual inspection of reliable patterns. Very few studies reported stage distributions of their baseline samples, preventing comparison of mean effect between studies intervening on pre-action or comprehensive stage-distribution samples.

Effect of Demographic Moderators

Some differences were found for demographic variables. Percent female was negatively related to effect size for dietary fat reduction, fruit and vegetable intake, and physical activity criteria. This most likely arises due to the fact that women in general engage in more health-conscious behavior than men and therefore have less to learn from the type of interventions provided here, suggesting a ceiling effect. Retention rate was a significant positive predictor of effect size for mammography and smoking cessation, suggesting that keeping participants involved in the study remains a vital component of intervention. Lack of finding significant moderators does not necessarily mean they are not present. Given small sample sizes, statistical power is low to detect these relationships. The database created for this study enables investigation of statistical power of these predictions in further follow-up studies.

Limitations

The main limitation of this study involves the wide differences among the studies in question. Tailoring is a relatively nascent field open to interpretation and various modalities of intervention. Messages differ in terms of writing style, language,

layout, amount of tailoring, behavior intervened upon, and assessment time points. Such disparity may limit the ability to compare studies in some instances.

Meta-analytic methods also carry limitations. Meta-analyses are correct regarding direction of effect about 80% of the time (Naylor, 1997), an acceptable but not perfect statistic. Statistically, meta-analysis places emphasis on variance among study effect sizes. With even a fair number of studies to compute a mean effect, power is limited to detect and predict between-study variability. Multivariate techniques require sample sizes (in this case number of studies) much larger than the number of predictors, a case that rarely exists for meta-analysis, a situation that limits modeling and discovering more specific conclusions. Since effect sizes appear similar across studies, combining effects across behaviors is theoretically justifiable. Such combination would increase the power to detect moderators, which was a significant limitation of the present analysis, preventing additional conclusions to be drawn regarding an optimal tailoring formula

The process of meta-analysis can also bias results. Generalizability may be limited if a limited sample of studies is found. Confounding of substantive and methodological features also occurs. If a difference appears in two groups that are also measured differently, the source of the discrepancy cannot be determined (Kazdin & Weisz, 1998). In addition, the nature of this study did not permit use of an additional coder to facilitate inter-rater reliability comparisons. This study did not calculate a methodology quality variable, which may be a valuable moderator in future analyses.

Sampling and publication bias inevitably skew the results of a meta-analysis. Searches are not able to locate all relevant studies, even with concerted effort. Despite

the intensive search for studies in the present analysis, a recently published work contains at least five studies not included here. Even with an intense search, the field has well-documented the publication bias problem such that significant studies are more often published than non-significant studies. This results in upward bias of the mean effect size. Since this study assessed outcomes largely from well-controlled and funded trials, publication bias may be limited.

Further questions

Differences between modality of intervention in terms of live counselor, print only, print plus counselor call, interactive terminal or web-based, or email reminder, whether tailored or retailored, could not be determined in this analysis. The majority of studies employed print tailored interventions alone, preventing meaningful comparison among these modalities. There does appear, however, to be added benefit of counselor calls for three of four behaviors, but this intervention is short-lived and surpassed by less expensive print retailored feedback over long-term follow up.

Dissemination remains a problem for many of these interventions since they require significant infrastructure to design and implement. Even if grant funding pays for initial development, interventions need to be continually administered to continue producing their effects on health behavior. Maintenance is a vital component in assessing the effectiveness of an intervention according to the RE-AIM framework suggested by Glasgow et al. (2002). An intervention with a large effect size will have little impact if it is not put into practice on a consistent basis. Intervening on a few thousand people will not reduce disease burden at a measurable population level. Traditional public health practice, usually at the state or local level, does not have

funding or organizational capacity in place to implement these programs. This leaves smaller, private organizations with an interest in prevention such as large employers or health insurance companies to implement these interventions on a population of their members. The cost/benefit ratio for long-term prevention must be clearly specified for such investments to be made. Unfortunately, prevention often makes little fiscal sense in the current health care setting. Beth Israel Medical Center in Manhattan, for example, developed a highly successful diabetes prevention and management protocol, which was subsequently halted due to lack of income from adverse diabetes sequelae such as amputations (Urbana, 2006).

Additional development of the methodology of tailored interventions would also facilitate their dissemination. These interventions require significant participant burden in terms of assessment to guide tailored feedback. The increasing use of electronic medical record technology enables information to be gathered on health behaviors and risk factors. Theoretical variables such as decisional balance can only be tailored through assessment, but much behavior and risk feedback information can be gathered directly from medical records without assessment burden. Many interventions in the present study employed a combination of targeted and tailored methodology by locating and targeting people most at risk for an outcome through medical record data and presenting them with a tailored intervention.

The survey of tailoring that this study enabled also revealed methodological flaws both in carrying out and reporting of outcomes. Meta-analysis largely relies on coding from written reports of each study, whether published or not. If meta-analysis is to guide progression of a field, outcome papers must report detailed (yet succinct)

accounts of their interventions. With the increasing use and usefulness of meta-analysis writers should report statistical results in formats that enable inclusion in a meta-analytic review. This would entail at least reporting means, standard deviations, confidence intervals, and actual p-values for every comparison. Stating results were “not significant” without reporting statistical data inordinately restricts the meta-analyst.

Greater detail is also needed in specifying how tailoring is accomplished in each particular study. Are constructs reported to have been included assessed at each timepoint to provide iterative feedback? How are constructs defined? What are their measurement properties? What do feedback reports look like? It becomes difficult to disentangle variables necessary to arrive at optimal tailoring methods without accurate reporting of the methods used. The present study suggests tailoring facilitates behavior change, but also highlights a troubling degree of uncertainty as to the definition of this term.

This study demonstrates that tailored interventions have the potential to impact health behaviors to a significant extent. The current database provides a solid foundation for assessing the effect of this methodology. New studies can easily be added, increasing sample size and enabling discernment of factors that could lead to more optimal tailoring.

APPENDICES

Appendix A: Coding Scheme

Variable Name	Instructions / Description
<i>Study Descriptors</i>	
Study Name	First author's last name and year of publication. In case of more than one study by same first author, second author is added.
Subgroup within study	In cases where outcomes are reported for subgroups such as women over 50 or adjusted versus unadjusted effect sizes, this subgroup is indicated.
Year	Enter the 4-digit year of publication. For manuscripts in preparation, enter year of draft.
Comparison	Names the comparison represented by the effect size difference (e.g. assessment only versus tailored).
Outcome	Describe outcome measured such as 24 hr quit or percent reaching Action or Maintenance.
Timepoint	Enter time of assessment for each effect size.
Country	List the country in which the study took place.
Significance	Enter report significance if provided for each effect size.
Multiple Behavior Study	Describe if intervention took place within context of simultaneous multiple behavior intervention study. 1 = single behavior 2 = two behaviors 3 = three behaviors 4 = four behaviors
# of interventions	Quantify number of times participants received feedback.
Stage of Sample	Describe if study participants were in all stages of change or if only preaction. 1 = preaction only 2 = all stages
Recruitment Strategy	Define if recruitment used proactive or reactive strategy 1 = proactive 2 = reactive
Recruitment Rate	Enter recruitment rate if reported. Usually applies only to proactively recruited samples. (Scaled from .01-1.0)
Retention Rate	Enter retention rate at each timepoint if available. (Scaled from .01-1.0).
Type of Analysis	Describe if statistical analysis employed intent to treat or all subjects remaining 1 = ITT 2 = All subjects remaining
Random Sampling	Code for recruitment proceeded using a random sample of participants 1 = random sampling 2 = convenience sample
Intervention Method	Describe method study used to intervene with participants 1 = Print 2 = Interactive Computer 3 = Telephone Counseling

	4 = Print + Telephone Counseling
Behavior	Categorize behavior upon which each outcome is based. 1 = smoking 2 = diet 3 = mammography screening 4 = physical activity
Delivery Site	Categorize location at which intervention was provided. 1 = home 2 = clinic 3 = worksite
Recruitment Site	Describe outlet through which participants were contacted 1 = RDD 2 = Call in center 3 = Clinic 4 = HMO membership 5 = school 6 = worksite 7 = store 8 = media fliers, announcements, etc 9 = church
Stage	Enter if intervention presented feedback on variable 1 = yes 2 = no
Decisional Balance	Enter if intervention presented feedback on variable 1 = yes 2 = no
Self-Efficacy	Enter if intervention presented feedback on variable 1 = yes 2 = no
Processes	Enter if intervention presented feedback on variable 1 = yes 2 = no
Strategies	Enter if intervention presented feedback on variable 1 = yes 2 = no
Social Support	Enter if intervention presented feedback on variable 1 = yes 2 = no
Motives	Enter if intervention presented feedback on variable 1 = yes 2 = no
Benefits	Enter if intervention presented feedback on variable 1 = yes 2 = no
Barriers	Enter if intervention presented feedback on variable 1 = yes 2 = no
Affect	Enter if intervention presented feedback on variable 1 = yes 2 = no
Beliefs	Enter if intervention presented feedback on variable 1 = yes 2 = no
Feedback	Enter if intervention presented feedback on variable 1 = yes

	2 = no
Goal setting	Enter if intervention presented feedback on variable 1 = yes 2 = no
Knowledge	Enter if intervention presented feedback on variable 1 = yes 2 = no
Risk	Enter if intervention presented feedback on variable 1 = yes 2 = no
Culture	Enter if intervention presented feedback on variable 1 = yes 2 = no
Addiction Level	Enter if intervention presented feedback on variable 1 = yes 2 = no
Theory	Enter theory upon which authors structured intervention. 1 = TTM 2 = HBM 3 = Social Cognitive 4 = Theory of Planned Behavior 5 = Attributional Theory 6 = TTM & Social Cognitive 7 = TTM & HBM
Initial N	Enter integer value of overall sample size at start of study
Age	Enter age of participants eligible for study 1 = 18+ 2 = 18-24 3 = 18-75 4 = 50+ etc
Mean Age	Enter reported mean age of participants
Study group	Enter main research center or study group that conceptualized the intervention.
Country Category	Code the country into one of the following: 1= United States 2= Non- US
Language	Enter the language in which the study was conducted (i.e., the language of the measures used).
Percent Female	Enter reported percent of female participants (Scale .01-1.0)
Percent Minority	Enter reported percent of non-white participants (Scale .01-1.0)
Publication	Enter the publication type and/or status of the study/data being used.
Publication Category	Code Publication description into the following categories: 1= technical manual 2= dissertation 3= peer review journal 4= manuscript in prep 5= masters thesis 6= unpublished data 7=conference presentation 8= research competency 9 = other

Appendix B: Studies Excluded from Analysis

Behavior	Citation	Exclusion Criteria
Diet	Blalock et al., 2002	Insufficient reporting
Diet	Brinberg & Axelson, 1990	Counselor based
Diet	Brinberg, Axelson, & Price, 2000	Counselor based
Diet	Brug & Assema, 2000	Results reported in (Brug, 1998)
Diet	Brug, Glanz, & Kok, 1997	No intervention provided
Diet	De Bourdeaudhuij, Brug, Vandelanotte, & Van Oost, 2002	Randomization and analysis according to family, not individual
Diet	Glanz, Murphy, Moylan, Evensen, & Curb, 2006	No control group
Diet	Glasgow, Toobert, Hampson, & Strycker, 2002	Counselor based
Diet	Jantz, Anderson, & Gould, 2002	Behavior change not reported
Diet	Oenema & Brug, 2003	Behavior change not reported
Diet	Oenema, Brug, & Lechner, 2001	Behavior change not measured
Diet	Sorensen et al., 1996	Community intervention, no tailored component
Diet	Winett et al., 1991	Not theoretically tailored, N = 30
Diet	Kreuter, Bull, Clark, & Oswald, 1999	Behavior change not measured
Diet	Tate, Wing, & Winett, 2001	Employed hand-tailored feedback
Diet	Tate, Jackvony, & Wing, 2003	Employed hand-tailored feedback
Diet Phys Act	Glasgow, Boles, McKay, Feil, & Barrera, 2003	Counselor based
Diet Physical Act	Clark, Hampson, Avery, & Simpson, 2004	Counselor based
Mammography	Allen & Bazargan-Hejazi, 2005	Counselor based
Mammography	Champion et al., 2003	Counselor based
Mammography	Gustafson et al., 2005	No tailored feedback provided
Mammography	Jibaja-Weiss, Volk, Kingery, Smith, & Holcomb, 2003	Not theoretically tailored
Mammography	McCaul & Wold, 2002	Employed manually tailored intervention
Mammography	Meldrum et al., 1994	Targeted intervention, not theoretically tailored
Mammography	Rimer et al., 2001	Reported in (Rimer, 2002)
Mammography	Stoddard et al., 2002	Counselor based
Mammography	Williams-Pichota, Pizarro, Schneider, Mowad, & Salovey, 2005	Reported in (Latimer, 2005)
Mammography	Ryan, Skinner, Farrell, & Champion, 2001	No intervention provided
Physical Act	Brownson et al., 2005	Results confounded with multilevel community intervention
Physical Act	Castro, King, & Brassington, 2001	Focused on maintenance
Physical Act	Demark-Wahnefried et al., 2003	No results reported
Physical Act	Marcus et al., 1998b	Results reported in (Bock, 2001)
Physical Act	Purath, Miller, McCabe, & Wilbur, 2004	Counselor based

Behavior	Citation	Exclusion Criteria
Physical Act	Vandelanotte & Bourdeaudhuij, 2003	No results reported
Physical Act	Van Sluijs, Van Poppel, Twisk, Brug, & Van Mechelen, 2005	Counselor based
Smoking	Ausems, Mesters, van Breukelen, & De Vries, 2002; Ausems, Mesters, van Breukelen, & De Vries, 2004	Employed social intervention
Smoking	Becona & Vázquez, 2001	Not theoretically tailored
Smoking	Burling, Marotta, González, & Moltzen, 1989	Not behaviorally tailored
Smoking	Carpenter, Watson, Raffety, & Chabal, 2003	Intervention focused on provider training
Smoking	Chouinard & Robichaud-Ekstrand, 2005	Counselor based
Smoking	Cobb, Graham, Bock, Papandonatos, & Abrams, 2005	No control group
Smoking	Klesges et al., 2006	Counselor based
Smoking	Orleans et al., 1998	Counselor based
Smoking	Pallonen et al., 1998	No control group
Smoking	Shegog et al., 2005	Behavior change not reported
Smoking	Wang & Etter, 2004	No control group
Smoking	Webb, Simmons, & Brandon, 2005	Behavior change not reported
Smoking	Wiggers et al., 2005	Behavior change not reported
Alcohol Use	Butler, Chiauuzi, Bromberg, Budman, & Buono, 2003	Insufficient number of same behavior for comparison.
Alcohol Use	Kypri et al., 2004	Insufficient number of same behavior for comparison.
Injury Prevention	McDonald et al., 2005	Insufficient number of same behavior for comparison.
Injury Prevention	Nansel et al., 2002	Insufficient number of same behavior for comparison.
Organ Donation	Reubsact, Brug, Kitslaar, van Hooff, & van den Borne, 2004	Insufficient number of same behavior for comparison.
Pain	Nicholson, Nash, & Andrasik, 2005	Insufficient number of same behavior for comparison.
Pain	Wilkie et al., 2001	No intervention provided
Risk Perception	Kreuter & Strecher, 1995	Insufficient number of same behavior for comparison.
Risk Perception	Kreuter & Strecher, 1996	Insufficient number of same behavior for comparison.
Risk Perception	Emmons, 2004	Insufficient number of same behavior for comparison.
Cancer screening	Kreuter et al., 2004	Behavior change not reported
Cancer screening	Marcus et al., 2005	Insufficient number of same behavior for comparison.
Cancer screening	de Nooijer, Lechner, & de Vries, 2002	Behavior change not reported
Sexual Risk Prevention	Bellis, Grimley, & Alexander, 2002	No intervention provided

Behavior	Citation	Exclusion Criteria
Sexual Risk Prevention	Chesney et al., 2003	Counselor based
Sexual Risk Prevention	Scholes et al., 2003	Insufficient number of same behavior for comparison.
Stress reduction	Evers et al., 2006	Insufficient number of same behavior for comparison.
Sun Protection	Bernhardt, 2001	Insufficient number of same behavior for comparison.
Sun Protection	Hornung et al., 2000	Insufficient number of same behavior for comparison.
Medication Adherence	Johnson, Driskell, Johnson, Prochaska et al., 2006	Insufficient number of same behavior for comparison.

BIBLIOGRAPHY

Agency for Healthcare Research and Quality. (2002). *Screening for breast cancer*.

Rockville, MD: U.S. Department of Health and Human Services.

Ajzen, I. (1985). From intention to actions: A theory of planned behavior. In J. Kuhl,

& J. Beckman (Eds.), *Action-control: From cognition to behavior* (pp. 11-39).

Heidelberg: Springer.

Ajzen, I., & Madden, T. J. (1986). Prediction of goal-directed behavior: Attitudes,

intentions, and perceived behavioral control. *Journal of Experimental Social*

Psychology, 22, 453-474.

Allen, B. J., & Bazargan-Hejazi, S. (2005). Evaluating a tailored intervention to

increase screening mammography in an urban area. *Journal of the National*

Medical Association, 97(10), 1350-1360.

Anderson, E. S., Winett, R. A., Wojcik, J. R., Winett, S. G., & Bowden, T. (2001). A

computerized social cognitive intervention for nutrition behavior: Direct and

mediated effects on fat, fiber, fruits and vegetables, self-efficacy and outcome

expectations among food shoppers. *Annals of Behavioral Medicine*, 23(2), 88-

100.

Armitage, C. J., & Conner, M. (2001). Efficacy of a minimal intervention to reduce fat

intake. *Social Science and Medicine*, 52(10), 1517-1524.

- Ausems, M., Mesters, I., van Breukelen, G., & De Vries, H. (2002). Short-term effects of a randomized computer-based out-of-school smoking prevention trial aimed at elementary schoolchildren. *Preventive Medicine, 34*(6), 581-589.
- Ausems, M., Mesters, I., van Breukelen, G., & De Vries, H. (2004). Effects of in-school and tailored out-of-school smoking prevention among Dutch vocational school students. *Health Education Research, 19*(1), 51-63.
- Aveyard, P., Griffin, C., Lawrence, T., & Cheng, K. K. (2003). A controlled trial of an expert system and self-help manual intervention based on the stages of change versus standard self-help materials in smoking cessation. *Addiction, 98*(3), 345-354.
- Bastani, R., Maxwell, A. E., Bradford, C., Das, I. P., & Yan, K. X. (1999). Tailored risk notification for women with a family history of breast cancer. *Preventive Medicine, 29*(5), 355-364.
- Becker, M., & Rosenstock, I. (Eds.). (1984). *Health care and human behavior* [Compliance with medical advice] . London: Academic Press.
- Becoña, E., & Vázquez, F. L. (2001). Effectiveness of personalized written feedback through a mail intervention for smoking cessation: A randomized-controlled trial in spanish smokers. *Journal of Consulting and Clinical Psychology, 69*(1), 33-40.
- Begg, C. B. (1994). Publication bias. In H. Cooper, & L. Hedges (Eds.), *The Handbook of Research Synthesis*. (pp. 399-409). New York: Sage.

- Bellis, J. M., Grimley, D. M., & Alexander, L. R. (2002). Feasibility of a tailored intervention targeting STD-related behaviors. *American Journal of Health Behavior, 26*(5), 378-385.
- Bernhardt, J. M. (2001). Tailoring messages and design in a web-based skin cancer prevention intervention. *International Electronic Journal of Health Education, 4*, 290-297.
- Blalock, S. J., DeVellis, B. M., Patterson, C. C., Campbell, M. K., Orenstein, D. R., & Dooley, M. A. (2002). Effects of an osteoporosis prevention program incorporating tailored educational materials. *American Journal of Health Promotion, 16*(3), 146-156.
- Bock, B. C., Marcus, B. H., Pinto, B. M., & Forsyth, L. H. (2001). Maintenance of physical activity following an individualized motivationally tailored intervention. *Annals of Behavioral Medicine, 23*(2), 79-87.
- Borland, R., Balmford, J., & Hunt, D. (2004). The effectiveness of personally tailored computer-generated advice letters for smoking cessation. *Addiction, 99*(3), 369-377.
- Borland, R., Balmford, J., Segan, C., Livingston, P., & Owen, N. (2003). The effectiveness of personalized smoking cessation strategies for callers to a quitline service. *Addiction, 98*(6), 837-846.

- Brinberg, D., Axelson, M. L., & Price, S. (2000). Changing food knowledge, food choice, and dietary fiber consumption by using tailored messages. *Appetite*, 35(1), 35-43.
- Brinberg, D., & Axelson, M. L. (1990). Increasing the consumption of dietary fiber: A decision theory analysis. *Health Education Research - Special Issue: Nutrition education*, 5(4), 409-420.
- Brownson, R. C., Hagood, L., Lovegreen, S. L., Britton, B., Caito, N. M., Elliott, M. B., et al. (2005). A multilevel ecological approach to promoting walking in rural communities. *Preventive Medicine*, 41(5-6), 837-842.
- Brug, J., van Assema, P., & de Vries, J. (1996). The impact of computer-tailored nutrition intervention. *Preventive Medicine*, 25, 236-242.
- Brug, J., & Assema, P. (2000). Differences in use and impact of computer-tailored dietary fat-feedback according to stage of change and education. *Appetite*, 34(3), 285-293.
- Brug, J., Glanz, K., & Kok, G. (1997). The relationship between self-efficacy, attitudes, intake compared to others, consumption, and stages of change related to fruit and vegetables. *American Journal of Health Promotion*, 12(1), 25-30.
- Brug, J., Steenhuis, I., van Assema, P., Glanz, K., & DeVries, H. (1999). Computer-tailored nutrition education: Differences between two interventions. *Health Education Research*, 14(2), 249-256.

- Brug, J., Campbell, M., & van Assema, P. (1999). The application and impact of computer-generated personalized nutrition education: A review of the literature. *Patient Education and Counseling*, 36(2), 145-156.
- Brug, J., Conner, M., Harré, N., Kremers, S., McKellar, S., & Whitelaw, S. (2005). The transtheoretical model and stages of change: A critique. Observations by five commentators on the paper by Adams, J. and White, M. (2004) Why don't stage-based activity promotion interventions work? *Health Education Research*, 20(2), 244-258.
- Brug, J., Glanz, K., Van Assema, P., Kok, G., & Van Breukelen, G. J. P. (1998). The impact of computer-tailored feedback and iterative feedback on fat, fruit, and vegetable intake. *Health Education & Behavior*, 25(4), 517-531.
- Bull, F. C., Jamrozik, K., & Blanksby, B. A. (1999). Tailored advice on exercise: Does it make a difference? *American Journal of Preventive Medicine*, 16(3), 230-239.
- Bull, F. C., Kreuter, M. W., & Scharff, D. P. (1999). Effects of tailored, personalized and general health messages on physical activity. *Patient Education and Counseling*, 36(2), 181-192.
- Burling, T. A., Marotta, J., González, R., & Moltzen, J. O. (1989). Computerized smoking cessation program for the worksite: Treatment outcome and feasibility. *Journal of Consulting and Clinical Psychology*, 57(5), 619-622.

- Burnett, K. F., Magel, P. E., Harrington, S., & Taylor, C. B. (1989). Computer-assisted behavioral health counseling for high school students. *Journal of Counseling Psychology, 36*(1), 63-67.
- Butler, S. F., Chiauzzi, E., Bromberg, J. I., Budman, S. H., & Buono, D. P. (2003). Computer-assisted screening and intervention for alcohol problems in primary care. *Journal of Technology in Human Services, 21*(3), 1-19.
- Campbell, M. K., Carbone, E., Honess-Morreale, L., Heisler-Mackinnon, J., Demissie, S., & Farrell, D. (2004). Randomized trial of a tailored nutrition education CD-ROM program for women receiving food assistance. *Journal of Nutrition Education Behavior, 36*(2), 58-66.
- Campbell, M. K., Bernhardt, J. M., Waldmiller, M., Jackson, B., Potenziani, D., Weathers, B., et al. (1999). Varying the message source in computer-tailored nutrition education. *Patient Education and Counseling, 36*(2), 157-169.
- Campbell, M. K., DeVellis, B. M., Strecher, V. J., & Ammerman, A. S. (1994). Improving dietary behavior: The effectiveness of tailored messages in primary care settings. *American Journal of Public Health, 84*(5), 783-787.
- Campbell, M. K., Tessaro, I., DeVellis, B., Benedict, S., Kelsey, K., Belton, L., et al. (2002). Effects of a tailored health promotion program for female blue-collar workers: Health works for women. *Preventive Medicine, 34*(3), 313-323.

- Cardinal, B. J., & Sachs, M. L. (1995). Prospective analysis of stage-of-exercise movement following mail-delivered, self-instructional exercise packets. *American Journal of Health Promotion, 9*(6), 430-432.
- Carpenter, K. M., Watson, J. M., Raffety, B., & Chabal, C. (2003). Teaching brief interventions for smoking cessations via an interactive computer-based tutorial. *Journal of Health Psychology, 8*(1), 149-160.
- Castro, C. M., King, A. C., & Brassington, G. S. (2001). Telephone versus mail interventions for maintenance of physical activity in older adults. *Health Psychology, 20*(6), 438-444.
- CDC. (2002). Annual smoking-attributable mortality, years of potential life lost, and economic costs - United States, 1995-1999. *Morbidity and Mortality Weekly Report, 51*(14), 300-303.
- CDC. (2004). Trends in intake of energy and macronutrients --- United States, 1971--2000. *Morbidity and Mortality Weekly Report, 53*(4), 80-82.
- CDC. (2005). Trends in leisure-time physical inactivity by age, sex, and race/ethnicity --- United States, 1994--2004. *Morbidity and Mortality Weekly Report, 54*(39), 991-994.
- CDC. (2006). Tobacco use among adults - United States, 2005. *Morbidity and Mortality Weekly Report, 55*(42), 1145-1148.

- CDC. (2007). Fruit and vegetable consumption among adults --- United States, 2005. *Morbidity and Mortality Weekly Report*, 56(10), 213-217.
- Centers for Disease Control. (2005). *The burden of chronic diseases and their risk factors: National and state perspectives 2004*. Atlanta: U.S. Department of Health and Human Services.
- Champion, V., Skinner, C. S., Hui, S., Monahan, P., Juliar, B., Daggy, J., et al. (2007). The effect of telephone versus print tailoring for mammography adherence. *Patient Education and Counseling*, 65(3), 416-423.
- Champion, V. L., Skinner, C. S., Menon, U., Seshadri, R., Anzalone, D. C., & Rawl, S. M. (2002). Comparisons of tailored mammography interventions at two months postintervention. *Annals of Behavioral Medicine*, 24(3), 211-218.
- Champion, V., Maraj, M., Hui, S., Perkins, A. J., Tierney, W., Menon, U., et al. (2003). Comparison of tailored interventions to increase mammography screening in nonadherent older women. *Preventive Medicine*, 36(2), 150-158.
- Chesney, M. A., Koblin, B. A., Barresi, P. J., Husnik, M. J., Celum, C. L., Colfax, G., et al. (2003). An individually tailored intervention for HIV prevention: Baseline data from the EXPLORE study. *American Journal of Public Health*, 93(6), 933-938.
- Chouinard, M., & Robichaud-Ekstrand, S. (2005). The effectiveness of a nursing inpatient smoking cessation program in individuals with cardiovascular disease. *Nursing Research*, 54(4), 243-254.

- Clark, M., Hampson, S. E., Avery, L., & Simpson, R. (2004). Effects of a tailored lifestyle self-management intervention in patients with type 2 diabetes. *British Journal of Health Psychology*, 9(3), 365-379.
- Clark, M. A., Rakowski, W., Ehrich, B., Rimer, B. K., Velicer, W. F., Dube, C. E., et al. (2002). The effect of a stage-matched and tailored intervention on repeat mammography. *American Journal of Preventive Medicine*, 22(1), 1-7.
- Cobb, N. K., Graham, A. L., Bock, B. C., Papandonatos, G., & Abrams, D. B. (2005). Initial evaluation of a real-world internet smoking cessation system. *Nicotine & Tobacco Research*, 7(2), 207-216.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. (2nd. ed.). Hillsdale, NJ: Erlbaum.
- Cohn, L. D., & Becker, B. J. (2003). How meta-analysis increases statistical power. *Psychological Methods*, 8(3), 243-253.
- Curry, S. J., McBride, C., Grothaus, L. C., Louie, D., & Wagner, E. H. (1995). A randomized trial of self-help materials, personalized feedback, and telephone counseling with nonvolunteer smokers. *Journal of Consulting and Clinical Psychology*, 63(6), 1005-1014.
- Curry, S. J., Wagner, E. H., & Grothaus, L. C. (1991). Evaluation of intrinsic and extrinsic motivation interventions with a self-help smoking cessation program. *Journal of Consulting and Clinical Psychology*, 59(2), 318-324.

- De Bourdeaudhuij, I., Brug, J., Vandelanotte, C., & Van Oost, P. (2002). Differences in impact between a family- versus an individual-based tailored intervention to reduce fat intake. *Health Education Research*, 17(4), 435-449.
- de Nooijer, J., Lechner, L., & de Vries, H. (2002). Tailored versus general information on early detection of cancer: A comparison of the reactions of dutch adults and the impact on attitudes and behaviors. *Health Education Research*, 17(2), 239-252.
- Delichatsios, H. K., Friedman, R. H., Glanz, K., Tennstedt, S., Smigelski, C., Pinto, B. M., et al. (2001). Randomized trial of a "talking computer" to improve adults' eating habits. *American Journal of Health Promotion*, 15(4), 215-224.
- Demark-Wahnefried, W., Clipp, E. C., McBride, C., Lobach, D. F., Lipkus, I., Peterson, B., et al. (2003). Design of FRESH START: A randomized trial of exercise and diet among cancer survivors. *Medicine & Science in Sports & Exercise*, 35(3), 415-424.
- Dijkstra, A. (2005). Working mechanisms of computer-tailored health education: Evidence from smoking cessation. *Health Education Research*, 20(5), 527-539.
- Dijkstra, A., De Vries, H., & Roijackers, J. (1999). Targeting smokers with low readiness to change with tailored and nontailored self-help materials. *Preventive Medicine: An International Journal Devoted to Practice and Theory*, 28(2), 203-211.

- Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2), 455-463.
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315, 629-634.
- Elder, J. P., Ayala, G. X., Campbell, N. R., Slymen, D., Lopez-Madurga, E. T., Engelberg, M., et al. (2005). Interpersonal and print nutrition communication for a Spanish-dominant Latino population: Secretos de la buena vida. *Health Psychology*, 24(1), 49-57.
- Emmons, K. M. (2004). Tailored computer-based cancer risk communication: Correcting colorectal cancer risk perception. *Journal of Health Communication*, 9(2), 127-141.
- Etter, J. (2005). Comparing the efficacy of two internet-based, computer-tailored smoking cessation programs: A randomized trial. *Journal of Medical Internet Research*, 7(1).
- Etter, J., & Perneger, T. V. (2004). Post-intervention effect of a computer tailored smoking cessation programme. *Journal of Epidemiology & Community Health*, 58(10), 849-851.
- Evers, K. E., Prochaska, J. O., Johnson, J. L., Mauriello, L. M., Padula, J. A., & Prochaska, J. M. (2006). A randomized clinical trial of a population- and

- transtheoretical model-based stress-management intervention. *Health Psychology*, 25(4), 521-529.
- Glanz, K., Murphy, S., Moylan, J., Evensen, D., & Curb, J. D. (2006). Improving dietary self-monitoring and adherence with hand-held computers: A pilot study. *American Journal of Health Promotion*, 20(3), 165-170.
- Glasgow, R. E., Boles, S. M., McKay, H. G., Feil, E. G., & Barrera, M. J. (2003). The D-net diabetes self-management program: Long-term implementation, outcomes, and generalization results. *Preventive Medicine: An International Journal Devoted to Practice and Theory*, 36(4), 410-419.
- Glasgow, R. E., Klesges, L. M., Dzewaltowski, D. A., Estabrooks, P. A., & Vogt, T. M. (2006). Evaluating the impact of health promotion programs: Using the RE-AIM framework to form summary measures for decision making involving complex issues. *Health Education Research*, 21(5), 688-694.
- Glasgow, R. E., Toobert, D. J., Hampson, S. E., & Strycker, L. A. (2002). Implementation, generalization and long-term results of the "choosing well" diabetes self-management intervention. *Patient Education and Counseling*, 48(2), 115-122.
- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5, 3-8.
- Goldfried, M. R. (1980). Toward the delineation of therapeutic change principles. *American Psychologist*, 35(11), 991-999.

- Gould, S. M., & Anderson, J. (2000). Using interactive multimedia nutrition education to reach low-income persons: An effectiveness evaluation. *Journal of Nutrition Education, 32*(4), 204-213.
- Greaney, M. L., Riebe, D., Garber, C. E., Rossi, J., Lees, F. D., Burbank, C. R., et al. (2007). *Long-term effects of a stage-based physical activity intervention for older adults: Results from the SENIOR project*. Unpublished manuscript.
- Gustafson, D. H., McTavish, F., Stengle, W., Ballard, D., Hawkins, R., Shaw, B. R., et al. (2005). Use and impact of eHealth system by low-income women with breast cancer. *Journal of Health Communication, 10*(Suppl1), 195-218.
- Hall, S. M., Delucchi, K. L., Velicer, W. F., Kahler, C. W., Ranger-Moore, J., Hedeker, D., et al. (2001). Statistical analysis of randomized trials in tobacco treatment: Longitudinal designs with dichotomous outcome. *Nicotine & Tobacco Research, 3*(3), 193-202.
- Harwell, M. (1997). An empirical study of hedge's homogeneity test. *Psychological Methods, 2*(2), 219-231.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego: Academic Press.
- Hedges, L. V., Cooper, H., & Bushman, B. J. (1992). Testing the null hypothesis in meta-analysis: A comparison of combined probability and confidence interval procedures. *Psychological Bulletin, 111*(1), 188-194.

- Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological methods*, 3(4), 486-504.
- Heimendinger, J., O'Neill, C., Marcus, A. C., Wolfe, P., Julesburg, K., Morra, M., et al. (2005). Multiple tailored messages are effective in increasing fruit and vegetable consumption among callers to the cancer information service. *Journal of Health Communication*, 10(Suppl1), 65-82.
- Heller, R. F., & Page, J. (2002). A population perspective to evidence based medicine: "Evidence for population health". *Journal of Epidemiology and Community Health*, 56(1), 45-47.
- Hornung, R. L., Lennon, P. A., Garrett, J. M., DeVellis, R. F., Weinberg, P. D., & Strecher, V. J. (2000). Interactive computer technology for skin cancer prevention targeting children. *American Journal of Preventive Medicine*, 18(1), 69-76.
- Huffcutt, A. I., & Arthur, W. (1995). Development of a new outlier statistic for meta-analytic data. *Journal of Applied Psychology*, 80(2), 327-334.
- Hunter, J., & Schmidt, F. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Beverly Hills, CA: Sage.
- Hunter, J. E., & Schmidt, F. L. (2000). Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Assessment*, 8(4), 275-292.
- Institute of Medicine. (2001). *Health and behavior*. National Academies Press.

- Jacobs, A. D., Ammerman, A. S., Ennett, S. T., Campbell, M. K., Tawney, K. W., Aytur, S. A., et al. (2004). Effects of a tailored follow-up intervention on health behaviors, beliefs, and attitudes. *Journal of Women's Health, 13*(5), 557-568.
- Jantz, C., Anderson, J., & Gould, S. M. (2002). Using computer-based assessments to evaluate interactive multimedia nutrition education among low-income predominantly hispanic participants. *Journal of Nutrition and Education Behavior, 34*(5), 252-260.
- Jeffrey, R. (1989). Risk behaviors and health: Contrasting individual and population perspectives. *American Psychologist, 44*(9), 1194-1202.
- Jibaja-Weiss, M. L., Volk, R. J., Kingery, P., Smith, Q. W., & Holcomb, J. D. (2003). Tailored messages for breast and cervical cancer screening of low-income and minority women using medical records data. *Patient Education & Counseling, 50*(2), 123-132.
- Jibaja-Weiss, M. L., Volk, R. J., Smith, Q. W., Holcomb, J. D., & Kingery, P. (2005). Differential effects of messages for breast and cervical cancer screening. *Journal of Health Care for the Poor and Underserved, 16*(1), 42-52.
- Johnson, S. S., Driskell, M. M., Johnson, J. L., Dymont, S. J., Prochaska, J. O., Prochaska, J. M., et al. (2006). Transtheoretical model intervention for adherence to lipid-lowering drugs. *Disease Management, 9*(2), 102-114.

- Johnson, S. S., Driskell, M. M., Johnson, J. L., Prochaska, J. M., Zwick, W., & Prochaska, J. O. (2006). Efficacy of a transtheoretical model-based expert system for antihypertensive adherence. *Disease Management*, 9(5), 291-301.
- Jones, H., & Rossi, J. (2003). Changes in diabetes self-care behaviors make a difference in glycemic control. *Diabetes Care*, 26(3), 732-737.
- Kazdin, A. E., & Weisz, J. R. (1998). Identifying and developing empirically supported child and adolescent treatments. *Journal of Consulting and Clinical Psychology*, 66(1), 19-36.
- King, D. K., Estabrooks, P. A., Strycker, L. A., Toobert, D. J., Bull, S. S., & Glasgow, R. E. (2006). Outcomes of a multifaceted physical activity regimen as part of a diabetes self-management intervention. *Annals of Behavioral Medicine*, 31(2), 128-137.
- Klie, P. (1987). Ninety Years of Psychometrics. In R. A. Peel (Ed.), *Essays in the History of Eugenics*. London: Galton Institute.
- Klesges, R. C., DeBon, M., Vander Weg, M. W., Haddock, C. K., Lando, H. A., Relyea, G. E., et al. (2006). Efficacy of a tailored tobacco control program on long-term use in a population of U.S. military troops. *Journal of consulting and clinical psychology*, 74(2), 295-306.
- Kopta, S. M., Lueger, R. J., Saunders, S. M., & Howard, K. I. (1999). Individual psychotherapy outcome and process research: Challenges leading to greater turmoil or a positive transition? *Annual Review of Psychology*, 50, 441-469.

- Kosma, M., Cardinal, B. J., & McCubbin, J. A. (2005). A pilot study of a web-based physical activity motivational program for adults with physical disabilities. *Disability and Rehabilitation*, 27(23), 1435-1442.
- Kreuter, M., Farrell, D., Olevitch, L., & Brennan, L. (2000). *Tailoring health messages: Customizing communication with computer technology*. Mahwah, NJ: Erlbaum.
- Kreuter, M. W., Bull, F. C., Clark, E. M., & Oswald, D. L. (1999). Understanding how people process health information: A comparison of tailored and nontailored weight-loss materials. *Health Psychology*, 18(5), 487-494.
- Kreuter, M. W., Skinner, C. S., Steger-May, K., Holt, C. L., Bucholtz, D. C., Clark, E. M., et al. (2004). Responses to behaviorally vs culturally tailored cancer communication among African American women. *American Journal of Health Behavior*, 28(3), 195-197.
- Kreuter, M. W., & Strecher, V. J. (1995). Changing inaccurate perceptions of health risk: Results from a randomized trial. *Health Psychology*, 14(1), 56-63.
- Kreuter, M. W., & Strecher, V. J. (1996). Do tailored behavior change messages enhance the effectiveness of health risk appraisal? Results from a randomized trial. *Health Education Research*, 11(1), 97-105.
- Kreuter, M. W., Sugg-Skinner, C., Holt, C. L., Clark, E. M., Haire-Joshu, D., Fu, Q., et al. (2005). Cultural tailoring for mammography and fruit and vegetable intake

among low-income African-American women in urban public health centers.

Preventive Medicine, 41(1), 53-62.

Kristal, A. R., Curry, S. J., Shattuck, A. L., Feng, Z., & Li, S. (2000). A randomized trial of a tailored, self-help dietary intervention: The Puget Sound eating patterns study. *Preventive Medicine*, 31(4), 380-389.

Kroeze, W., Werkman, A., & Brug, J. (2006). A systematic review of randomized trials on the effectiveness of computer-tailored education on physical activity and dietary behaviors. *Annals of Behavioral Medicine*, 31(3), 205-223.

Kypri, K., Saunders, J. B., Williams, S. M., McGee, R. O., Langley, J. D., Cashell-Smith, M. L., et al. (2004). Web-based screening and brief intervention for hazardous drinking: A double-blind randomized controlled trial. *Addiction*, 99(11), 1410-1417.

Latimer, A. E., Katulak, N. A., Mowad, L., & Salovey, P. (2005). Motivating cancer prevention and early detection behaviors using psychologically tailored messages. *Journal of Health Communication*, 10(Suppl1), 137-155.

Laygo, R., O'Donohue, W., Hall, S., Kaplan, A., Wood, R., Cummings, J., et al. (2003). Preliminary results from the Hawaii integrated healthcare project II. In Cummings, N. A., and O'Donohue, W. (Eds.). *Behavioral Health as Primary Care: Beyond Efficacy to Effectiveness*. Reno: Context Press.

Lennox, A. S., Osman, L. M., Reiter, E., Robertson, R., Friend, J., McCann, I., et al. (2001). Cost effectiveness of computer tailored an non-tailored smoking cessation

- letters in general practice: Randomised controlled trial. *British Medical Journal*, 322(7299), 1396-1400.
- Lipkus, I. M., Rimer, B. K., Halabi, S., & Strigo, T. S. (2000). Can tailored interventions increase mammography use among HMO women? *American Journal of Preventive Medicine*, 18(1), 1-10.
- Lipkus, I. M., Lyna, P. R., & Rimer, B. K. (1999). Using tailored interventions to enhance smoking cessation among African-Americans at a community health center. *Nicotine & Tobacco Research*, 1(1), 77-85.
- Lipsey, M., & Wilson, D. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*, 48, 1181-1209.
- Lipsey, M., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- Llewelyn, S., & Kennedy, P. (2003). *Handbook of Clinical Health Psychology*. Chichester, UK: Wiley.
- Lutz, S. F., Ammerman, A. S., Atwood, J. R., Campbell, M. K., DeVellis, R. F., & Rosamond, W. D. (1999). Innovative newsletter interventions improve fruit and vegetable consumption in healthy adults. *Journal of the American Dietetic Association*, 99(6), 705-709.

- Marcus, A. C., Mason, M., Wolfe, P., Rimer, B. K., Lipkus, I., Strecher, V., et al. (2005). The efficacy of tailored print materials in promoting colorectal cancer screening: Results from a randomized trial involving callers to the national cancer institute's cancer information service. *Journal of Health Communication, 10*(Suppl1), 83-104.
- Marcus, B. H., Emmons, K. M., Simkin-Silverman, L. R., Linnan, L. A., Taylor, E. R., Bock, B. C., et al. (1998a). Evaluation of motivationally tailored vs. standard self-help physical activity interventions at the workplace. *American Journal of Health Promotion, 12*(4), 246-253.
- Marcus, B. H., Bock, B. C., Pinto, B. M., Forsyth, L. H., Roberts, M. B., & Traficante, R. M. (1998b). Efficacy of an individualized, motivationally-tailored physical activity intervention. *Annals of Behavioral Medicine, 20*(3), 174-180.
- Marsh, K. L., Johnson, B. T., & Carey, M. P. (2001). Conducting meta-analyses of HIV prevention literatures from a theory-testing perspective. *Evaluation & the Health Professions, 24*(3), 255-276.
- McCaul, K. D., & Wold, K. S. (2002). The effects of mailed reminders and tailored messages on mammography screening. *Journal of Community Health, 27*(3), 181-190.
- McDonald, E. M., Solomon, B., Shields, W., Serwint, J. R., Jacobsen, H., Weaver, N. L., et al. (2005). Evaluation of kiosk-based tailoring to promote household safety

- behaviors in an urban pediatric primary care practice. *Patient Education and Counseling*, 58(2), 168-181.
- Meldrum, P., Turnbull, D., Dobson, H. M., Colquhoun, C., Gilmour, W. H., & McIlwaine, G. M. (1994). Tailored written invitations for second round breast cancer screening: A randomised controlled trial. *Journal of Medical Screening*, 1(4), 245-248.
- Nansel, T. R., Weaver, N., Donlin, M., Jacobsen, H., Kreuter, M. W., & Simons-Morton, B. (2002). Baby, be safe: The effect of tailored communications for pediatric injury prevention provided in a primary care setting. *Patient Education & Counseling*, 46(3), 175-190.
- Napolitano, M. A., Fotheringham, M., Tate, D., Sciamanna, C., Leslie, E., Owen, N., et al. (2003). Evaluation of an internet-based physical activity intervention: A preliminary investigation. *Annals of Behavioral Medicine*, 25(2), 92-99.
- Naylor, C. D. (1997). Meta-analysis and the meta-epidemiology of clinical research. *British Medical Journal*, 315(7109), 617-619.
- Nicholson, R., Nash, J., & Andrasik, F. (2005). A self-administered behavioral intervention using tailored messages for migraine. *Headache*, 45(9), 1124-1139.
- Noar, S. M., & Zimmerman, R. S. (2005). Health behavior theory and cumulative knowledge regarding health behaviors: Are we moving in the right direction? *Health Education Research*, 20(3), 275-290.

- Norman, P., & Conner, M. (1995). The theory of planned behavior and exercise: An investigation into the role of prior behavior, behavioral intentions, and attitude variability. *European Journal of Social Psychology*, 25, 403-415.
- Oenema, A., Brug, J., & Lechner, L. (2001). Web-based tailored nutrition education: Results of a randomized controlled trial. *Health Education Research*, 16(6), 647-660.
- Oenema, A., & Brug, J. (2003). Feedback strategies to raise awareness of personal dietary intake: Results of a randomized controlled trial. *Preventive Medicine*, 36(4), 429-439.
- Oenema, A., Tan, F., & Brug, J. (2005). Short-term efficacy of a web-based computer-tailored nutrition intervention: Main effects and mediators. *Annals of Behavioral Medicine*, 29(1), 54-63.
- Ogden, J. (2000). *Health Psychology: A Textbook*. Buckingham, UK: Open University Press.
- Orleans, C. T., Boyd, N. R., Bingler, R., Sutton, C., Fairclough, D., Heller, D., et al. (1998). A self-help intervention for african american smokers: Tailoring cancer information service counseling for a special population. *Preventive Medicine*, 193-1997, 27(5, Pt.2), S61-S70.
- Orwin, R. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics*, 8, 157-159.

- Overton, R. C. (1998). A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods*, 3(3), 354-379.
- Owen, N., Ewins, A., & Lee, C. (1989). Smoking cessation by mail: A comparison of standard and personalized correspondence course formats. *Addictive Behaviors*, 14(4), 355-363.
- Pallonen, U. E., Velicer, W. F., Prochaska, J. O., Rossi, J. S., Bellis, J. M., Tsoh, J. Y., et al. (1998). Computer-based smoking cessation interventions in adolescents: Description, feasibility, and six-month follow-up findings. *Substance Use & Misuse*, 33(4), 935-965.
- Patrick, K., Sallis, J. F., Prochaska, J. J., Lydston, D. D., Calfas, K. J., Zabinski, M. F., et al. (2001). A multicomponent program for nutrition and physical activity change in primary care: PACE+ for adolescents. *Archives of Pediatric and Adolescent Medicine*, 155(8), 940-946.
- Peters, K., & Elster, A. (2002). *Roadmaps for Clinical practice: A Primer on Population-Based Medicine*. AMA Press.
- Peterson, T. R., & Aldana, S. G. (1999). Improving exercise behavior: An application of the stages of change model in a worksite setting. *American Journal of Health Promotion*, 13(4), 229-232.
- Petty, K., & Elster, A. (1981). *Attitudes and persuasion: Classical and contemporary approaches*. Dubuque, IA: Brown.

- Pham, B., Platt, R., McAuley, L., Sampson, M., Klassen, T. P., & Moher, D. (2001). Detecting and minimizing the impact of for publication bias: An empirical study of methods. *Evaluation and the Health Professions*, 24(2), 109-125.
- Pinto, B., Friedman, R., Marcus, B. H., Kelley, H., Tennstedt, S., & Gillman, M. W. (2002). Effects of a computer-based, telephone-counseling system on physical activity. *American Journal of Preventive Medicine*, 23(2), 113-120.
- Poisal, J. A., Truffer, C., Smith, S., Sisko, A., Cowan, C., Keehan, S., et al. (2007). Health spending projections through 2016: Modest changes obscure part D's impact. *Health Affairs*, 26(2), 242-253.
- Prochaska, J. O., & DiClemente, C. (1982). Transtheoretical therapy: Toward a more integrative model of change. *Psychotherapy: Theory Research and Practice*, 19, 276-288.
- Prochaska, J. O., & Velicer, W. F. (2004). Integrating population smoking cessation policies and programs. *Public Health Reports*, 119(3), 244-252.
- Prochaska, J. O., Velicer, W. F., Fava, J. L., Ruggiero, L., Laforge, R. G., Rossi, J. S., et al. (2001). Counselor and stimulus control enhancements of a stage-matched expert system intervention for smokers in a managed care setting. *Preventive Medicine*, 32(1), 23-32.
- Prochaska, J. O. (1999). How do people change, and how can we change to help many more people? In M.A. Hubble, B.L. Duncan & S.D. Miller (Eds.), *The Heart and Soul of Change*. Washington, DC: American Psychological Association Press.

Prochaska, J. O., DiClemente, C. C., Velicer, W. F., & Rossi, J. S. (1993).

Standardized, individualized, interactive, and personalized self-help programs for smoking cessation. *Health Psychology, 12*(5), 399-405.

Prochaska, J. O., Velicer, W. F., Fava, J. L., Rossi, J. S., & Tsoh, J. Y. (2001).

Evaluating a population-based recruitment approach and a stage-based expert system intervention for smoking cessation. *Addictive Behaviors, 26*(4), 583-602.

Prochaska, J. O., Velicer, W. F., Redding, C., Rossi, J. S., Goldstein, M., DePue, J., et al. (2005). Stage-based expert systems to guide a population of primary care patients to quit smoking, eat healthier, prevent skin cancer, and receive regular mammograms. *Preventive Medicine, 41*(2), 406-416.

Prochaska, J. O., Velicer, W. F., Rossi, J. S., Redding, C. A., Greene, G. W., Rossi, S. R., et al. (2004). Multiple risk expert systems interventions: Impact of simultaneous stage-matched expert system interventions for smoking, high-fat diet, and sun exposure in a population of parents. *Health Psychology, 23*(5), 503-516.

Prochaska, J. J., & Sallis, J. F. (2004). A randomized controlled trial of single versus multiple health behavior change: Promoting physical activity and nutrition among adolescents. *Health Psychology, 23*(3), 314-318.

Purath, J., Miller, A. M., McCabe, G., & Wilbur, J. (2004). A brief intervention to increase physical activity in sedentary working women. *Canadian Journal of Nursing Research, 36*(1), 76-91.

- Rakowski, W., Ehrich, B., Goldstein, M. G., Rimer, B. K., Pearlman, D. N., Clark, M. A., et al. (1998). Increasing mammography among women aged 40-74 by use of a stage-matched, tailored intervention. *Preventive Medicine, 27*(5, Pt.1), 748-756.
- Rakowski, W., Lipkus, I. M., Clark, M. A., Rimer, B. K., Ehrich, B., Lyna, P. R., et al. (2003). Reminder letter, tailored stepped-care, and self-choice comparison for repeat mammography. *American Journal of Preventive Medicine, 25*(4), 308-314.
- Reubsaet, A., Brug, J., Kitslaar, J., van Hooff, J. P., & van den Borne, H. W. (2004). The impact and evaluation of two school-based interventions on intention to register an organ donation preference. *Health Education Research, 19*(4), 447-456.
- Rimer, B. K., Orleans, C. T., Fleisher, L., & Cristinzio, S. (1994). Does tailoring matter? the impact of a tailored guide on ratings and short-term smoking-related outcomes for older smokers. *Health Education Research, 9*(1), 69-84.
- Rimer, B. K., Halabi, S., Skinner, C. S., Kaplan, E. B., Crawford, Y., Samsa, G. P., et al. (2001). The short-term impact of tailored mammography decision-making interventions. *Patient Education and Counseling, 43*(3), 271-287.
- Rimer, B. K., Halabi, S., Skinner, C. S., Lipkus, I. M., Strigo, T. S., Kaplan, E. B., et al. (2002). Effects of a mammography decision-making intervention at 12 and 24 months. *American Journal of Preventive Medicine, 22*(4), 247-257.
- Rosenstock, I. (1966). Why people use health services. *Millbank Memorial Fund Quarterly, 44*, 94-124.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results.

Psychological Bulletin, 86(3), 638-641.

Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin*, 118(2), 183-192.

Rossi, J. S. (2003). Comparison of the use of significance testing and effect sizes in theory-based health promotion research. *43rd Annual Meeting of the Society for Multivariate Experimental Psychology*, Keystone, CO.

Ryan, G. L., Skinner, C. S., Farrell, D., & Champion, V. L. (2001). Examining the boundaries of tailoring: The utility of tailoring versus targeting mammography interventions for two distinct populations. *Health Education Research*, 16(5), 555-556.

Rychetnik, L., Frommer, M., Hawe, P., & Shiell, A. (2002). Criteria for evaluating evidence on public health interventions. *Journal of Epidemiological Community Health*, 56(2), 119-127.

Saul Shiffman, Jean A. Paty, Jeffrey M. Rohay, Michael E. Di Marino, & Joe Gitchell. (2000). The efficacy of computer-tailored smoking cessation material as a supplement to nicotine polacrilex gum therapy. *Archives of Internal Medicine*, 160, 1675-1681.

Saywell, R. M. J., Champion, V. L., Skinner, C. S., Menon, U., & Daggy, J. (2004). A cost-effectiveness comparison of three tailored interventions to increase mammography screening. *Journal of Women's Health*, 13(8), 909-918.

- Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods*, 1(2), 115-129.
- Scholes, D., McBride, C. M., Grothaus, L., Civic, D., Ichikawa, L. E., Fish, L. J., et al. (2003). A tailored minimal self-help intervention to promote condom use in young women: Results from a randomized trial. *AIDS*, 17(10), 1547-1556.
- Shadish, W. R. (1996). Meta-analysis and the exploration of causal mediating processes: A primer of examples, methods and issues. *Psychological Methods*, 1, 47-56.
- Shegog, R., McAlister, A. L., Hu, S., Ford, K. C., Meshack, A. F., & Peters, R. J. (2005). Use of interactive health communication to affect smoking intentions in middle school students: A pilot test of the "headbutt" risk assessment program. *American Journal of Health Promotion*, 19(5), 334-338.
- Shiffman, S., Paty, J. A., Rohay, J. M., Di Marino, M. E., & Gitchell, J. G. (2001). The efficacy of computer-tailored smoking cessation material as a supplement to nicotine patch therapy. *Drug and alcohol dependence*, 64(1), 35-46.
- Skinner, C. S., Strecher, V. J., & Hospers, H. (1994). Physicians' recommendations for mammography: Do tailored messages make a difference? *American Journal of Public Health*, 84(1), 43-49.

- Skinner, C. S., Campbell, M. K., Rimer, B. K., Curry, S., & Prochaska, J. O. (1999). How effective is tailored print communication? *Annals of Behavioral Medicine*, 21(4), 290-298.
- Sorensen, G., Thompson, B., Glanz, K., Feng, Z., Kinne, S., DiClemente, C., et al. (1996). Work site-based cancer prevention: Primary results from the working well trial. *American Journal of Public Health*, 86(7), 939-947.
- Stern, J., & Simes, R. (1997). Publication bias: Evidence of delayed publication in a cohort study of clinical research projects. *British Medical Journal*, 315, 640-645.
- Sterne, J. A., Gavaghan, D., & Egger, M. (2000). Publication and related bias in meta-analysis: Power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology*, 53(11), 1119-1129.
- Stevens, V. J., Glasgow, R. E., Toobert, D. J., Karanja, N., & Smith, K. S. (2003). One-year results from a brief, computer-assisted intervention to decrease consumption of fat and increase consumption of fruits and vegetables. *Preventive Medicine*, 36(5), 594-600.
- Stoddard, A. M., Fox, S. A., Costanza, M. E., Lane, D. S., Andersen, M. R., Urban, N., et al. (2002). Effectiveness of telephone counseling for mammography: Results from five randomized trials. *Preventive Medicine*, 34(1), 90-99.
- Strecher, V. J. (1999). Computer-tailored smoking cessation materials: A review and discussion. *Patient Education and Counseling*, 36(2), 107-117.

- Strecher, V. J., Kreuter, M., Den Boer, D., & Kobrin, S. (1994). The effects of computer-tailored smoking cessation messages in family practice settings. *Journal of Family Practice*, 39(3), 262-270.
- Strecher, V. J., Marcus, A., Bishop, K., Fleisher, L., Stengle, W., Levinson, A., et al. (2005). A randomized controlled trial of multiple tailored messages for smoking cessation among callers to the cancer information service. *Journal of Health Communication*, 10(Suppl1), 105-118.
- Strecher, V. J., Shiffman, S., & West, R. (2005). Randomized controlled trial of a web-based computer-tailored smoking cessation program as a supplement to nicotine patch therapy. *Addiction*, 100(5), 682-688.
- Tate, D. F., Jackvony, E. H., & Wing, R. R. (2003). Effects of internet behavioral counseling on weight loss in adults at risk for type 2 diabetes: A randomized trial. *Journal of the American Medical Association*, 289(14), 1833-1836.
- Tate, D. F., Wing, R. R., & Winett, R. A. (2001). Using internet technology to deliver a behavioral weight loss program. *Journal of the American Medical Association*, 285(9), 1172-1177.
- US Department of Health and Human Services. (2001). *The Surgeon General's Call to Action to Prevent and Decrease Overweight and Obesity* (No. 017-001-00551-7). Washington, DC: U.S. Government Printing Office.
- Urbana, I. (Jan 1, 2006). In the treatment of diabetes, success often does not pay. *New York Times*.

- Van Sluijs, E. M. F., Van Poppel, M. N. M., Twisk, J. W. R., Brug, J., & Van Mechelen, W. (2005). The positive effect on determinants of physical activity of a tailored, general practice-based physical activity intervention. *Health Education Research*, 20(3), 345-356.
- Vandelandotte, C., & Bourdeaudhuij, I. D. (2003). Acceptability and feasibility of a computer-tailored physical activity intervention using stages of change: Project FAITH. *Health Education Research*, 18(3), 304-317.
- Vandelandotte, C., De Bourdeaudhuij, I., Sallis, J. F., Spittaels, H., & Brug, J. (2005). Efficacy of sequential or simultaneous interactive computer-tailored interventions for increasing physical activity and decreasing fat intake. *Annals of Behavioral Medicine*, 29(2), 138-146.
- Velicer, W. F., Prochaska, J. O., Fava, J. L., Rossi, J. S., Redding, C. A., Laforge, R. G., et al. (2000). Using the transtheoretical model for population-based approaches to health promotion and disease prevention. *Homeostasis in Health and Disease*, 40(5), 174-195.
- Velicer, W. F., Friedman, R. H., Fava, J. L., Gulliver, S. B., Keller, S., Sun, X., et al. (2006). Evaluating nicotine replacement therapy and stage-based therapies in a population-based effectiveness trial. *Journal of Consulting and Clinical Psychology*, 74(6), 1162-1172.
- Velicer, W. F., & Prochaska, J. O. (2004). A comparison of four self-report smoking cessation outcome measures. *Addictive Behaviors*, 29(1), 51-60.

Velicer, W. F., Prochaska, J. O., Fava, J. L., Laforge, R. G., & Rossi, J. S. (1999).

Interactive versus noninteractive interventions and dose-response relationships for stage-matched smoking cessation programs in a managed care setting. *Health Psychology, 18*(1), 21-28.

Wang, J., & Etter, J. (2004). Administering an effective health intervention for smoking cessation online: The international users of stop-tabac. *Preventive Medicine, 39*(5), 962-968.

Webb, M. S., Simmons, V. N., & Brandon, T. H. (2005). Tailored interventions for motivating smoking cessation: Using placebo tailoring to examine the influence of expectancies and personalization. *Health Psychology, 24*(2), 179-188.

Weinstein, N. D. (1984). Why it won't happen to me: Perceptions of risk factors and susceptibility. *Health Psychology, 3*, 431-457.

Wiggers, L. C. W., Oort, F. J., Dijkstra, A., de Haes, J. C. J. M., Legemate, D. A., & Smets, E. M. A. (2005). Cognitive changes in cardiovascular patients following a tailored behavioral smoking cessation intervention. *Preventive Medicine, 40*(6), 812-821.

Wilkie, D. J., Huang, H., Berry, D. L., Schwartz, A., Lin, Y., Ko, N., et al. (2001). Cancer symptom control: Feasibility of a tailored, interactive computerized program for patients. *Family & Community Health, 24*(3), 48-62.

Williams-Piehota, P., Pizarro, J., Schneider, T. R., Mowad, L., & Salovey, P. (2005). Matching health messages to monitor-blunter coping styles to motivate screening mammography. *Health Psychology, 24*(1), 58-67.

Winett, R. A., Wagner, J. L., Moore, J. F., Walker, W. B., Hite, L. A., Leahy, M., et al. (1991). An experimental evaluation of a prototype public access nutrition information system for supermarkets. *Health Psychology, 10*(1), 75-78.