

2016

Extending Theory-Based Quantitative Predictions to New Health Behaviors

Leslie Ann D. Brick
University of Rhode Island

Wayne Velicer
University of Rhode Island, wvelicer@uri.edu

See next page for additional authors

Follow this and additional works at: http://digitalcommons.uri.edu/cprc_facpubs

**The University of Rhode Island Faculty have made this article openly available.
Please let us know how Open Access to this research benefits you.**

This is a pre-publication author manuscript of the final, published article.

Terms of Use

This article is made available under the terms and conditions applicable towards Open Access Policy Articles, as set forth in our [Terms of Use](#).

Citation/Publisher Attribution


Brick, L.A. D., Velicer, W. F., Redding, C. A., & Rossi, J. S. Extending Theory-Based Quantitative Predictions to New Health Behaviors. *International Journal of Behavioral Medicine*, 23(2), 123-134.
Available at: <http://dx.doi.org/10.1007/s12529-015-9506-y>

This Article is brought to you for free and open access by the Cancer Prevention Research Center at DigitalCommons@URI. It has been accepted for inclusion in Cancer Prevention Research Center Faculty Publications by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons@etal.uri.edu.

Authors

Leslie Ann D. Brick, Wayne Velicer, Colleen A. Redding, and Joseph S. Rossi

Extending Theory-Based Quantitative Predictions to New Health Behaviors

Leslie Ann D. Brick¹  · Wayne F. Velicer¹ · Colleen A. Redding¹ · Joseph S. Rossi¹ · James O. Prochaska¹

© International Society of Behavioral Medicine 2015

Abstract

Background Traditional null hypothesis significance testing suffers many limitations and is poorly adapted to theory testing.

Purpose A proposed alternative approach, called *Testing Theory-based Quantitative Predictions*, uses effect size estimates and confidence intervals to directly test predictions based on theory.

Method This paper replicates findings from previous smoking studies and extends the approach to diet and sun protection behaviors using baseline data from a Transtheoretical Model behavioral intervention ($N=5407$). Effect size predictions were developed using two methods: (1) applying refined effect size estimates from previous smoking research or (2) using predictions developed by an expert panel.

Results Thirteen of 15 predictions were confirmed for smoking. For diet, 7 of 14 predictions were confirmed using smoking predictions and 6 of 16 using expert panel predictions. For sun protection, 3 of 11 predictions were confirmed using smoking predictions and 5 of 19 using expert panel predictions.

Conclusion Expert panel predictions and smoking-based predictions poorly predicted effect sizes for diet and sun protection constructs. Future studies should aim to use previous empirical data to generate predictions whenever possible. The best results occur when there have been several iterations of predictions for a behavior, such as with smoking, demonstrating that expected values begin to converge on the

population effect size. Overall, the study supports necessity in strengthening and revising theory with empirical data.

Keywords Effect sizes · Health behavior change · Confidence intervals · Null hypothesis significance testing · Transtheoretical model

Introduction

Theory development is essential for the generation and support of research ideas for scientific advancement. Traditional null hypothesis significance testing (NHST) has been the modus operandi for testing research questions across many branches of science since the early 1900s [1]. Modern day NHST is the result of a blending of two schools of thought: the Fisherian approach, which simply features a statistical test of the null hypothesis, and the Neyman–Pearson approach, which introduces the alternative hypothesis, a fixed alpha level, specification of one- or two-tailed regions, as well as type I and II errors [1, 2]. The institutionalization of this blended approach has been attributed to confusion surrounding hypothesis testing, resulting in ritualistic thinking about statistical methods [3]. For example, the focus of a statistical test under the NHST framework considers the rejection of a null hypothesis based on a conditional probability of the data, given that the null hypothesis is true (e.g., a p value). Consequently, this framework does not provide support for the theory or for the alternative hypothesis [1, 4]. In fact, the theory has little impact on the formulation of the statistical test and a lack of power can lead to inaccurate conclusions. This limited focus on the null hypothesis has been criticized for almost as long as significance testing has existed, though it persists as a common ritual in the social sciences [2, 3]. Furthermore, rejection of a null hypothesis provides no information on the magnitude

✉ Leslie Ann D. Brick
leslie.brick@gmail.com

¹ Cancer Prevention Research Center, University of Rhode Island, Kingston, RI, UK

of a difference and is greatly affected by the size of the sample, the alpha level, and the effect size. Such reliance on p value cutoffs, which themselves are arbitrary, can lead to misunderstanding of results and misinterpretation of conclusions [1, 2, 5]. Therefore, the limitations of NHST warrant the movement toward more relevant and rigorous approaches to theory testing.

One such approach, called *Testing Theory-based Quantitative Predictions* (TTQP), uses effect size indices and confidence intervals (CI) to directly test quantitative predictions generated by prior knowledge and theory [6]. In this approach, a researcher specifies a numeric prediction of an expected effect and compares it to the effect size in the sample data. In this way, essential information regarding the magnitude and direction of the observed effect is demonstrated. Further, CIs surrounding these estimates provide a means of “testing” the numeric prediction: if the CI contains the predicted value, the prediction is confirmed and if the predicted value falls outside of the CI, the prediction is not confirmed and explanations for failed predictions are examined. In this manner, TTQP shifts the focus of a study toward examination of prior knowledge or theory (e.g., in a form of quantitative start values) rather than a traditional null (or nil) hypothesis.

The use of TTQP is advantageous because it emphasizes estimation of effect sizes surrounded by confidence intervals, which can be employed for intervention development and subsequent meta-analysis, among other applications [5]. This orientation facilitates comparisons across studies, while direct tests of theory or previous estimates leads to more refined values as the number of studies accumulates. Further, it is especially useful in the development of new behavioral interventions in that the TTQP approach can be used to guide decision-making towards the most influential psychological constructs to be employed in an intervention. For example, self-efficacy and social support may have statistically significant effects on quitting smoking, but self-efficacy may be more important earlier in the behavior change process, thereby potentially serving as a better focus during early stages of an intervention. Accordingly, TTQP can be used to determine or rank which constructs result in greater effects at certain points during the behavior change process to guide decision making for better tailored, optimized, and more effective interventions [7, 8].

The TTQP approach is best applied in the context of an explicit theory that provides a framework to test predictions regarding the relationship between behavioral constructs or variables. Previous research has employed the TTQP approach using the Transtheoretical Model (TTM) of health behavior change. The TTM was developed as an integration of numerous theories of psychotherapy to describe the process of intentional behavior change [9–11]. Initial studies utilizing TTM theory were conducted using a comparative analysis of smokers and self-changers, but it was rapidly expanded to

numerous other health behaviors including alcohol and substance abuse, obesity, medication compliance, bullying, and mammography screening [12]. The theoretical framework of the TTM regards behavior change as the temporal movement through a series of five ordered stages, ranging from not ready to change a behavior to maintaining a changed behavior. These stages include: Precontemplation (PC), Contemplation (C), Preparation (PR), Action (A), and Maintenance (M). Moreover, the TTM integrates multiple mediating constructs (i.e., the Pros and Cons from the Decisional Balance (DB) and Self-Efficacy (SE) scales) and a number of other constructs (i.e., the Processes of Change (POC)) beyond the stages of change, which represent activities or experiences that individuals engage in during the modification of behavior. These additional variables have been found to predict successful change [12].

The constructs of the TTM are specific to each behavior and provide important guides for intervention, as individuals across each stage of change utilize them to differing extents at different times. For example, individuals in PC tend to weigh the Cons of changing their behavior higher than individuals in PR [13]. This difference can be characterized by an effect size that should be consistent across studies, as it represents a meaningful change central to TTM theory. In fact, meta-analyses have supported the strong and weak principles of progress across 48 behaviors [14]. The strong principle purports that progress from PC to A involves about one standard deviation increase in the Pros of changing. The weak principle purports that progress from PC to A involves about a half a standard deviation decrease in the Cons of changing. Consequently, the strong theoretical basis and the quantitative strengths of the TTM lend well for its use with TTQP approach.

In fact, previous research in theory testing using the TTM has contributed to the development of the TTQP approach. Velicer and colleagues [15] conducted a longitudinal smoking study that tested 40 differential a priori, theory-based predictions using sample data from a large intervention. These predictions consisted of explicit quantitative predictions regarding the expected effect size (e.g., omega-squared) of TTM constructs during the course of an intervention. However, the study used significance tests instead of CIs to support the hypotheses, falling victim to the weaknesses of NHST. Consequently, Velicer and colleagues [16] replicated the previous 40 predictions using a new sample of smokers using effect size predictions and CIs instead of significance tests. The study also included an analysis of prediction failure in order to promote thoughtful consideration of failed predictions as a way to improve theory. Most failures were attributed to sample fluctuation or the need to recalibrate and revise effect size estimates.

A cross-sectional study also formalized the TTQP approach in a sample of smokers [6]. In this study, 15 predictions

were made considering the magnitude of effect sizes for TTM constructs (Decisional Balance, Self-Efficacy, and POC) across the first three stages of change, thereby, representing participants who have not yet changed their behavior. Hypotheses were generated to predict the magnitude of effect size for a given construct, such as the Pros, across the stages of change. Eleven of 15 predictions were confirmed by sample data. Missed predictions were determined to be a result of four potential issues: sample fluctuation, a need for theory revision, theory incorrect, or a need for further calibration of effect size categories. Consideration of missed predictions is not intended to be a post hoc modification of original hypotheses, but rather a method to promote a culture of replication and the updating of theory based on empirical findings. It is discussed in greater detail below.

Current Study

The TTQP approach has been conducted using the framework of the TTM applied to smoking cessation [6, 16]. However, the approach needs to be replicated with new data and applied using previously unexamined health behaviors, such as diet and sun protection. The present study draws on secondary data from a TTM behavioral intervention and consists of three parts: *study one* presents a replication of findings from Velicer et al. [6] by using previous empirical data as predictions for estimation in a new sample; *study two* and *study three* extend TTQP methodology to estimation of effects for diet and sun protection, respectively. In the absence of previously existing results for diet and sun protection from which to base predictions, two methods for prediction development are considered: (1) a direct application of previously refined values based on smoking cessation and (2) the use of an expert panel to develop new predictions. The goal of these three studies is to provide a demonstration of how to apply and extend the TTQP approach using a strong theoretical framework guiding a set of statistical predictions.

Development and Testing of Predictions

Effect size predictions are conducted a priori and involve several steps. If previous empirical research exists, point estimates may be used as starting value for prediction. Study 1 in this work will apply previous findings from smoking research as predictions to new sample data. In the absence of previous data, as is the case for studies 2 and 3, theoretically based descriptions of the expected values are designated for each prediction. One critical issue discussed in this paper investigates two methods for developing initial predictions for previously unexamined behaviors. The first approach involves a direct application of formerly refined values from comparable theoretical constructs within another behavior paradigm. In this example, smoking cessation predictions

have been refined by two previous studies [6, 17] and can be used as predicted values for diet and sun protection behaviors. This approach can be applied to behaviors that are measured using similar framework; however, many health behaviors contain unique constructs that do not directly translate between behaviors. Therefore, the second method for developing initial predictions considers predictions generated by a panel of experts. This approach is intended for use as a starting point when no previous data exists to empirically inform predictions. Methodological and substantive experts involved in the study propose the most theoretically reasonable predictions. Verbal predictions begin as “small,” “medium,” and “large” and are subsequently translated into quantitative values based on Cohen’s [18] guidelines (i.e., a small effect equals a prediction of .01).

Both methods for developing predictions emphasize the iterative and replicative nature of the TTQP approach. Predictions should be viewed as starting points to inform the current study, ultimately updating the knowledge base of effects through replication with empirical data. The use of an expert panel should only be used as a starting point when no comparable research findings are available. In a way, the TTQP approach is conceptually similar to Bayesian inference, such that sample data is used to update prior knowledge to generate a new understanding. As with Bayesian approaches, thorough consideration of prior knowledge is essential for the success of the approach and the accuracy of the posterior distribution.

After predictions are generated, effect size estimates and CIs are generated from sample data. If the CI surrounding the observed effect contains the predicted value, then the prediction is confirmed. If the predicted value falls outside of the CI, then the prediction is not confirmed and explanations for failed predictions are examined. Examination of failed predictions is an important component, as it serves to inform future studies by investigating potential reasons for a missed prediction (see below, the “[Examination of Predictions](#)” section) that can help to guide theory improvement.

Examination of Predictions

After effect sizes and CIs have been generated, an investigation of the confirmations and misses is necessary to fully understand how the theory withstood the evidence provided by sample data. Careful thought into the potential reasons behind missed predictions moves researchers away from dichotomous accept/reject thinking and forces consideration of how to best revise theory in light of empirical evidence. This understanding is then used to update or support the theory for future studies.

Several explanations exist that may elucidate failed predictions. First, sample fluctuation plays a large role since the use of CI permits a small number of misses due to chance fluctuations. As the number of predictions increase, the chance of a

miss due to sample fluctuation increases as well. These misses tend to be “near misses” and are very close to falling within the interval generated by sample data, but instead fall just outside of it. Thus, some of these near misses may in fact be confirmed in a future study with another sample while others may indicate the need for slight adjustment of expected values. With all of the explanations for failed predictions, replication using independent samples is necessary to refine theory. Second, prediction revision may be required when an observation falls far away from the predicted effect. Such may be the case when a medium effect was predicted and a very large effect was obtained. In this case, the theory made an inadequate prediction and needs to be revised to instead predict large effects. Third, a prediction may be incorrect when observations are undoubtedly discordant with predictions such that the theory led to an overwhelmingly incorrect prediction. It may be the case that the theory itself needs major reconsideration to better reflect a construct, as opposed to a slight revision or a near miss. Finally, further recalibration may be needed when observations and their CIs do not align with any of the predicted values. It is noted that Cohen’s classifications of effect size are very broadly defined and were intended only as a guide to initial estimates. A small effect may be represented by a prediction of .02 rather than .01. In some cases, the creation of new categories may be needed to reflect effects if data reveal values outside of Cohen’s guidelines (i.e., such as those well beyond .14). In this way, the effect size classifications may be “recalibrated” to better reflect the population and content area [6].

As theorists become familiar with the effects found in a specific area of research or population, they should update the theoretical classifications of effect sizes to better reflect each construct. These “expected values” should be generated post hoc in order to serve as starting points for prediction in subsequent studies. The intention is not to modify original predictions but rather to update expectations regarding the magnitude of effect size estimates. Ideally, as theory predictions become more refined through replication, empirical studies serve to update quantitative values that best reflect effects that can be expected by the theory.

Methods

Study Design

A secondary analysis was conducted using baseline data from a population-based multiple risk factor behavioral intervention in the USA [19, 20]. Participants were recruited and assessed for smoking, diet, and sun protection behaviors. Smoking behavior was assessed in regard to self-reported daily smoking habits. Diet was assessed in regard to a high fat diet: greater than 30 % calories and total score on the Dietary

Behavior Questionnaire [21, 22]. Sun protection was assessed using self-reported exposure: 15 or more minutes of exposure per day or inconsistent SPF-15 use and total score on the Sun Protection Behavior Scale [23, 24]. Human subjects review boards of all participating institutions granted approval for this study.

Sample

A large health insurance organization provided patient information and 5,407 primary care patients agreed to take part in the study. Eligible participants were at risk for at least one of three behaviors (smoking, high fat diet, and inadequate sun protection); therefore, all participants in the study are characterized by one of the first three stages of change (PC, C, or PR) in order to examine effect sizes for measures within the sample of at risk participants. All data from the current study were drawn from baseline measurement. Participants were 68.0 % female, 96.7 % white, 1.3 % Hispanic, and had a mean age of 44.7 years (SD=12.7).

Measures

Three behaviors were examined in this study including smoking, diet, and sun protection. Core TTM constructs (e.g., stage, Decisional Balance, Self-Efficacy, and POC) were measured using items specific to each behavior and included a slightly different conceptual framework depending on the nature of the behavior. See below for descriptions of these measures.

Stages of Change Participants were classified into one of five stages of change using an algorithm that assessed their readiness to change. For diet and sun protection behaviors, stage was assessed based on an individual’s perception of their readiness to change, and then adjusted based on a series of questions regarding their habits and behaviors to best reflect their readiness to change. Detailed discussion of stage of change measures for smoking [25], for diet [21, 22, 26], and for sun protection [23, 24, 27] has been published previously.

Decisional Balance Cognitive and motivational aspects of decision-making are measured by the Decisional Balance Inventory [13, 28]. These two constructs are the Pros and Cons of engaging in a behavior. For smoking and diet, the Pros reflect perceived benefits of engaging in the unhealthy behavior and Cons reflect disadvantages. For sun protection, the reasoning is reversed. The Pros reflect benefits of sun protection while the Cons reflect the difficulties.

Self-Efficacy A person’s self-efficacy, or belief that they can prevent or cope with the temptation to fall back into unhealthy or high-risk behavior, is measured using measures of

confidence and situational temptations [29, 30]. The temptation to engage in negative health behavior in various situations was measured by the Situational Temptations Inventory [30]. For smoking, the subscales were positive/social (pos/soc), negative/affective (neg/aff), and habit/addictive (habit/add) aspects of the temptation to smoke. For dietary fat, subscales included: positive/social, negative/affective, and difficult situations [31]. Confidence was measured by a scale designed to assess confidence to engage in sun protection behavior. This scale has two subscales, sun avoidance confidence and sunscreen use confidence, and is represented by a sum score of each subscale, with a higher score representing higher confidence. Both subscales measured responses using a 5-point Likert scale (1=not at all confident to 5=extremely confident).

Processes of Change The processes of change represent ten different behavioral and experiential strategies for changing behavior [32]. Experiential processes include: consciousness raising (CR), dramatic relief (DR), environmental reevaluation (ER), self-reevaluation (SR), and social liberation (SO). Behavioral processes include stimulus control (SC), counter conditioning (CC), reinforcement management (RM), self-liberation (SL), and helping relationships (HR). These ten variables make up the “core” of the TTM processes of change, but additional processes specific to certain behaviors are often added to examine different concepts. For diet, one additional process was added, interpersonal systems control (IS), resulting in a total of 11 processes. For sun protection, six additional processes were added: health care provider (HC), interpersonal systems control (IS), reducing exposure (RE), regret (RG), sunscreen use (SS), and health responsibility (HT). Stimulus control (SC) was not included resulting in a total of 15 processes. The processes of behavior change represent ten different behavioral and cognitive and affective experiential strategies for changing behavior [32]. These processes help to explain how and when changes in cognition, emotion, and behavior take place as an individual moves through the stages of change. Cognitive and affective experiential processes include: consciousness raising (CR; getting the facts), dramatic relief (DR; paying attention to feelings), environmental reevaluation (ER; noticing your effect on others), self-reevaluation (SR; creating a new self-image), and social liberation (SO; noticing public support). Behavioral processes include stimulus control (SC; managing your environment), counter conditioning (CC; using substitutes for problem behaviors), reinforcement management (RM; using rewards), self-liberation (SL; making a commitment), and helping relationships (HR; getting support). These ten variables make up the “core” of the TTM processes of change, but additional processes specific to certain behaviors are often added to examine different concepts. For diet, one additional process, interpersonal systems control (IS), is included and

refers to the monitoring of people and situations that may trigger the problem behavior. Thus, diet behavior consists of a total of 11 processes. Sun protection behavior consists of 15 total processes, with nine of the ten original processes (stimulus control was excluded as it is generally not relevant for sun protective behavior) and six additional processes [33–35]. These include: interpersonal systems control (IS; as mentioned above), health care provider (HC; awareness of health care provider’s advice to protect skin from sun exposure), reducing exposure (RE; limiting exposure to sun by seeking shade or wearing clothes/hat during peak hours), regret (RG; feeling embarrassed or guilty when you get too much sun or sunburn), sunscreen use (SS; keeping and carrying sunscreen so that it is handy and ready to use when needed), and health responsibility (HT; realizing that protecting one’s skin from sun exposure is part of being a healthy adult).

Initial Effect Size Predictions

Effect size predictions for smoking have been examined previously [6]. Study 1 used previous findings as predictions in order to replicate and validate the observed effects for smoking behavior. In the absence of previous empirical values for diet and sun protection, two approaches were taken to generate predictions. The first approach directly applied the previously published smoking-based predictions [6]. Since these predictions came from smoking-based constructs, predictions unique to diet and sun protection were not generated. Therefore, a second approach was employed to provide predictions for these constructs. In this approach, a panel of seven TTM, diet, and sun protection experts met to discuss appropriate predictions for each psychological construct. All members of the panel had reviewed empirical results for TTM interventions involving the three behaviors across previous studies, but not in the form of the effect size estimates employed in this paper. Thus, these estimates were informed by empirical results over many studies and represent theoretically and empirically informed judgments. Consequently, these predictions are used as starting values for the iterative TTQP process.

A discussion of all predictions follows. First, study 1 replicated smoking findings using predictions from previous studies. Then, studies 2 and 3 used either these smoking-based values from study 1 or expert panel predictions.

Study 1: Smoking

Fifteen effect size predictions for smoking were generated. Predictions represent a direct replication of results integrating the recalibrated effect sizes from previous research [6]. The results from this study were used as smoking-based predictions for studies 2 and 3.

Study 2: Diet

Part 1: Smoking-Based Predictions Fourteen predictions for diet were generated from smoking-based predictions for constructs common to both behaviors. These include: Pros, Cons, positive/social, negative/affective, counter conditioning, consciousness raising, dramatic relief, environmental reevaluation, helping relations, reinforcement management, stimulus control, self-liberation, social liberation, and self-reevaluation. Since the difficult situations and interpersonal systems control measures do not exist for smoking behavior, no predictions from smoking data exist to provide starting values.

Part 2: Expert Panel-Based Predictions Since not all of the diet-based constructs were common with smoking behavior, an expert panel met to generate sixteen predictions for all constructs independent of the smoking-based predictions.

Study 3: Sun Protection

Part 1: Smoking-Based Predictions Eleven predictions for sun protection were generated from smoking-based predictions for constructs common to both behaviors. These include: Pros, Cons, counter conditioning, consciousness raising, dramatic relief, environmental reevaluation, helping relations, reinforcement management, self-liberation, social liberation, and self-reevaluation. Since the sunscreen confidence, sun avoidance confidence, health responsibility, interpersonal systems control, reducing exposure, regret, or health care provider measures do not exist for smoking behavior, no predictions from smoking data exist to provide starting values.

Part 2: Expert Panel-Based Predictions Since not all of the sun protection-based constructs were common with smoking behavior, an expert panel met to generate 19 predictions for all the constructs independent of the smoking-based predictions.

Statistical Analyses

One-way between-groups fixed effects analysis of variance were conducted in SAS 9.2 to compare each intervention construct with baseline stage classification for the behaviors examined in the three studies. Participants who entered the intervention were considered at risk for possible health risks (smoking, diet, and sun protection); therefore, the participants in the current study were limited to membership in the first three stages (PC, C, and PR) to examine the magnitude of effect sizes among early stages of change. Population-adjusted effect size, omega-squared (ω^2) was calculated to represent the variance in each construct that can be accounted for by the first three stages. Larger effect sizes demonstrate more behavioral change. CIs for ω^2

were calculated at the 99 % level [36] because of the large sample sizes.

Results

Stage distributions of each behavior for the sample used from the intervention are displayed in Table 1.

Study 1: Smoking

Thirteen of 15 predictions were confirmed for smoking behavior. Table 2 shows a summary of results for the 15 variables. The two misses include the Pros and SL variables. An effect size of zero was predicted for the Pros scale, but a value of $\omega^2=.022$ was observed with a lower limit of .004, indicating a near miss. For SL, a value of $\omega^2=.19$ was predicted, but a value of $\omega^2=.101$ was observed. Figure 1 displays predicted effect sizes and observed CIs for each construct. Misses are indicated by a dot falling outside of the upper or lower boundaries of the CI.

Examination of Prediction Misses It is evident that although the predicted effect size estimates for the two misses (i.e., for Pros and SL) were not confirmed, they fell just outside the observed 99 % CIs. The Pros scale was posited to have no effect across the first three stages, but was observed to have a small effect with lower limit of .004. This finding can be considered a near miss. The interval surrounding SL was wide, with an upper CI of .164, but was found to be lower than predicted. This finding may be due in part to study specific fluctuation, such that the observed effects for SL in this sample tended to be lower than what has been estimated in previous studies, or may indicate the need for slight decrease in the expected effect size for SL.

Study 2: Diet

Part 1: Smoking-Based Predictions Seven of 14 predictions were confirmed for diet using smoking-based predictions.

Table 1 Sample sizes^a of stage membership for smoking, diet, and sun protection behaviors

Stage	Smoking	Diet	Sun Protection
PC	388	2,012	1,236
C	578	514	932
PR	294	1,266	1,707

PC precontemplation, C contemplation, PR preparation

^a Total sample sizes for each behavior vary and are lower than the intervention total ($N=5,407$) because participants who were not at risk (i.e., in PC, C, and PR) for given a behavior were excluded from each specific analysis

Table 2 Quantitative predictions and observed estimates of ω^2 for smoking with 99 % CIs

Construct	Measure	<i>N</i>	ω^2_{pred}	ω^2_{obs}	L-CI	U-CI	Confirm?
Decisional balance	Pros	1230	0.000	0.022	0.004	0.047	No
	Cons	1223	0.070	0.054	0.025	0.090	Yes
Self-efficacy	Pos/Soc	1240	0.010	0.017	0.002	0.040	Yes
	Habit St.	1241	0.010	0.018	0.002	0.041	Yes
	Neg/Aff	1241	0.000	0.008	0.000	0.026	Yes
Processes	CC	586	0.050	0.030	0.002	0.074	Yes
	SL	588	0.190	0.101	0.046	0.164	No
	SO	585	0.010	0.000	0.000	0.020	Yes
	CR	588	0.090	0.057	0.016	0.111	Yes
	RM	587	0.030	0.010	0.000	0.042	Yes
	ER	589	0.040	0.017	0.000	0.053	Yes
	DR	588	0.090	0.099	0.044	0.161	Yes
	SR	589	0.180	0.183	0.113	0.254	Yes
	SC	588	0.070	0.043	0.008	0.092	Yes
	HR	581	0.020	0.004	0.000	0.030	Yes

ω^2_{pred} predicted effect size, ω^2_{obs} observed effect size, *L-CI* lower confidence interval, *U-CI* upper confidence interval, *Pos/Soc* positive/social, *Neg/Aff* negative/affective, *Dif Sit* difficult situations, *CC* counter conditioning, *SL* self-liberation, *SO* social liberation, *CR* consciousness raising, *RM* reinforcement management, *ER* environmental reevaluation, *DR* dramatic relief, *SR* self-reevaluation, *SC* stimulus control, *HR* helping relationships

Table 3 shows a summary of predictions and results for the 14 constructs. The misses include: Neg/Aff, CC, HR, RM, SC, SO, and SR. Figure 2 displays predicted effect sizes and observed CIs for constructs in increasing order of effect size.

Part 2: Expert Panel-Based Predictions Six of 16 predictions were confirmed for diet behavior based on expert panel judgment. Table 2 shows a summary of predictions and results for the 16 constructs. The misses include: Pros, Neg/Aff, CC, DR, ER, HR, IS, RM, SC, and SL. Figure 2 displays predicted

effect sizes and observed CIs for constructs in increasing order of effect size.

Examination of Misses Seven misses resulted from using smoking-based predictions and ten misses resulted from using expert panel judgment. Results indicate that both methods did poorly at predicting effect size estimates, with the expert panel doing considerably poorly. Four of the smoking-based predictions (i.e., HR, SC, SL, and SR) consisted of misses falling just outside of the CI bounds. These near misses indicate that the prediction was very close to the observed effect, requiring

Fig. 1 Comparison of predicted and observed ω^2 effect size values for smoking constructs surrounded by 99 % confidence intervals. *Pos/Soc* positive/social, *Neg/Aff* negative/affective, *Dif Sit* difficult situations, *CC* counter conditioning, *SL* self-liberation, *SO* social liberation, *CR* consciousness raising, *RM* reinforcement management, *ER* environmental re-evaluation, *DR* dramatic relief, *SR* self-re-evaluation, *SC* stimulus control, *HR* helping relationships

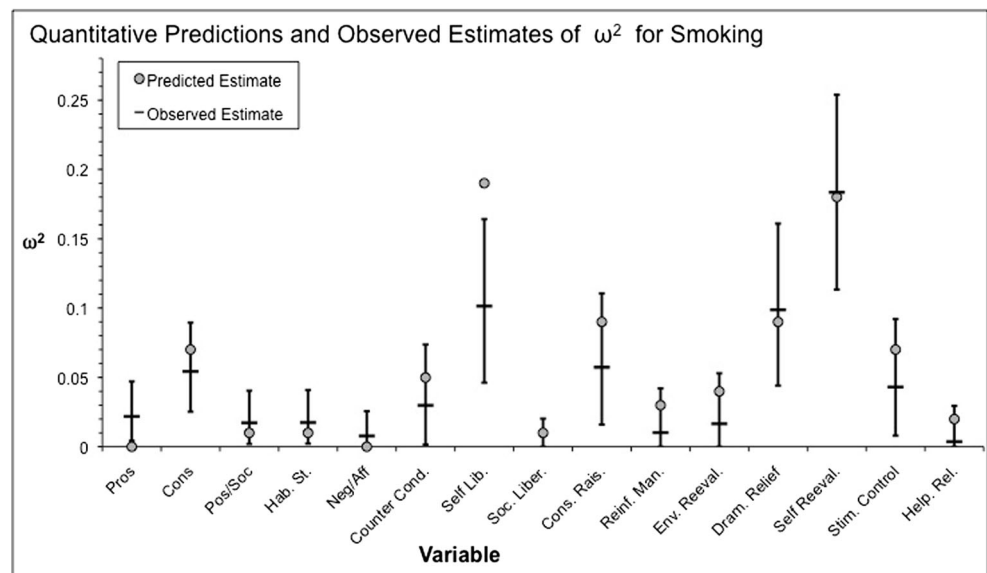


Table 3 Expert panel and theory-based predictions with observed estimates of ω^2 for diet with 99 % CIs

Construct	Measure	N	Expert panel		Smoking-based		Observed data		
			ω^2_{pred}	Confirm?	ω^2_{pred}	Confirm?	ω^2_{obs}	L-CI	U-CI
Decisional balance	Pros	3,444	0.00	No	0.022	Yes	0.018	0.008	0.030
	Cons	3,471	0.06	Yes	0.054	Yes	0.048	0.027	0.061
Self-efficacy	Pos/Soc	3,501	0.01	Yes	0.017	Yes	0.014	0.005	0.025
	Neg/Aff	3,508	0.01	No	0.008	No	0.057	0.039	0.078
	Dif Sit	3,450	0.01	Yes	–	–	0.019	0.009	0.033
Processes	CC	1,693	0.01	No	0.030	No	0.070	0.042	0.102
	CR	1,698	0.06	Yes	0.057	Yes	0.076	0.047	0.109
	DR	1,697	0.06	No	0.099	Yes	0.112	0.077	0.150
	ER	1,690	0.06	No	0.017	Yes	0.030	0.012	0.054
	HR	1,693	0.01	No	0.004	No	0.030	0.012	0.054
	IS	1,693	0.01	No	–	–	0.036	0.015	0.061
	RM	1,684	0.01	No	0.010	No	0.068	0.040	0.100
	SC	1,692	0.06	No	0.060	No	0.026	0.009	0.048
	SL	1,697	0.14	No	0.101	Yes	0.101	0.067	0.137
	SO	1,685	0.01	Yes	0.000	No	0.016	0.003	0.035
	SR	1,688	0.14	Yes	0.183	No	0.131	0.094	0.170

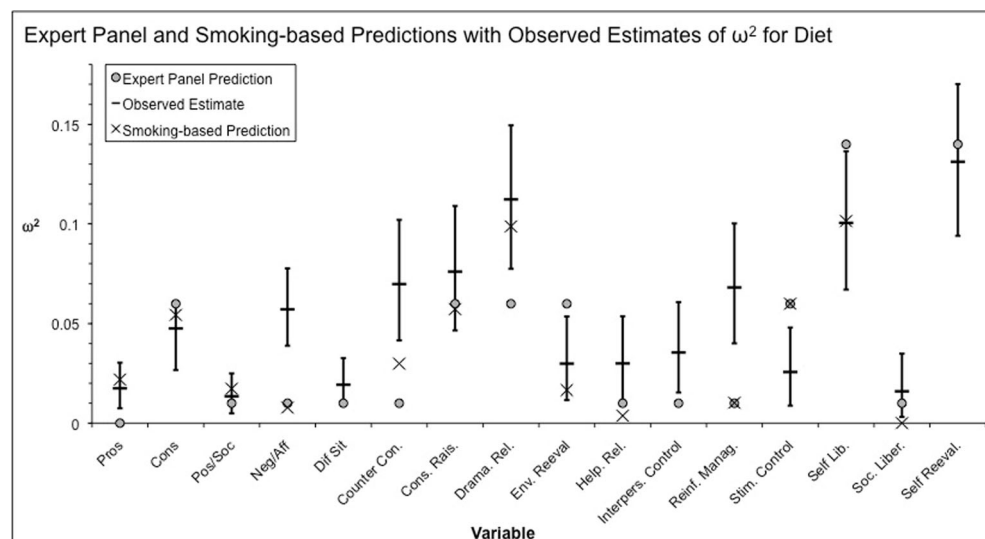
ω^2_{pred} predicted effect size, ω^2_{obs} observed effect size, *L-CI* lower confidence interval, *U-CI* upper confidence interval, *Pos/Soc* positive/social, *Neg/Aff* negative/affective, *Dif Sit* difficult situations, *CC* counter conditioning, *CR* consciousness raising, *DR* dramatic relief, *ER* environmental reevaluation, *HR* helping relationships, *IS* interpersonal systems control, *RM* reinforcement management, *SC* stimulus control, *SL* self-liberation, *SO* social liberation, *SR* self-reevaluation

very slight adjustment to expected values or perhaps reflecting small sample specific effects. Three of the misses (i.e., SS, Neg/Aff, and RM) require prediction revision, as they are clearly under or overestimating observed effects that appear to be medium sized effects.

In the second part of study 2, an expert panel correctly predicted six of 16 estimates. Six of the ten (i.e., Pros, ER, HR, IS, SC, and SL) failed predictions may be considered near misses, as they fell just outside of the CI. These near misses

indicate that the prediction was very close to the observed effect, requiring very slight adjustment to expected values or perhaps reflecting small sample specific effects. The four remaining failed predictions require prediction revision. In general, predictions tended to be low, thus larger than expected effects were observed for many variables. Small effects were predicted for three variables (i.e., Neg/Aff, CC, and RM), but medium effects were found. DR was predicted to have a medium effect, but was instead observed to be large.

Fig. 2 Comparison of expert panel and smoking-based predictions with observed ω^2 effect size estimates for diet constructs surrounded by 99 % confidence intervals. *Pos/Soc* positive/social, *Neg/Aff* negative/affective, *Dif Sit* difficult situations, *CC* counter conditioning, *CR* consciousness raising, *DR* dramatic relief, *ER* environmental re-evaluation, *HR* helping relationships, *IS* interpersonal systems control, *RM* reinforcement management, *SC* stimulus control, *SL* self-liberation, *SO* social liberation, *SR* self-re-evaluation



Study 3: Sun Protection

Part 1: Smoking-Based Predictions Three of 11 predictions were confirmed for sun protection using smoking-based predictions. Table 4 shows a summary of results for the 11 constructs. The misses include Pros, Cons, ER, HR, RM, SR, SL, and SO. Figure 3 displays predicted effect sizes and observed CIs for constructs in increasing order of effect size.

Part 2: Expert Panel-Based Predictions Five of 19 predictions were confirmed for sun protection based on expert panel judgment. Table 3 shows a summary of results for the 19 variables. The misses included Cons, sunscreen confidence, sun avoidance confidence, CC, ER, HR, HT, IS, RE, RM, RG, SR, and SS. Figure 3 displays predicted effect sizes and observed CIs for constructs in increasing order of effect size.

Examination of Misses Both the smoking-based predictions and the expert panel-based prediction performed extremely poorly when applied to prediction of sun protection effects. Smoking-based predictions generated eight misses. Three of these were near misses (i.e., Pros, HR, and SL), indicating that the prediction was very close to the observed effect and may

require very slight adjustment to expected values or may small sample specific effects. The remaining five misses (i.e., Cons, SR, SL, RM and ER) for smoking-based predictions required prediction revision. These effects were underestimated and expected values should be increased.

In the second part of study 3, the expert panel predictions incorrectly predicted 14 estimates. An examination of the misses reveals that many of them were discordant with theory, signifying the need for prediction revision and new effect size categories. First, three variables HR, SR, and HC can be interpreted as near misses. Next, slight prediction revision may improve expected values from another seven misses. ER, IS, and SS were all predicted to have a medium effect, but demonstrated large effects. SS and RM were predicted to have small effects, but demonstrated medium effects. Finally, RE was predicted to have a large effect, but demonstrated a medium effect.

The remaining four misses from the expert panel were greatly discordant from predictions. The prediction for Cons was more than one effect size class lower than the observed effect. It was predicted to have no effect, but was found to be a large effect. Many of the remaining observations were also discordant with theory but revealed extremely large effect

Table 4 Expert panel and theory-based predictions with observed estimates of ω^2 for sun protection with 99 % CIs

Construct	Measure	N	Expert panel		Smoking-based		Observed sata		
			ω^2_{pred}	Confirm?	ω^2_{pred}	Confirm?	ω^2_{obs}	L-CI	U-CI
Decisional balance	Pros	3,805	0.06	Yes	0.022	No	0.047	0.031	0.065
	Cons	3,799	0.00	No	0.054	No	0.170	0.143	0.197
Self-efficacy	Sunsc.	3,799	0.01	No	–	–	0.240	0.210	0.268
	Sun Av.	3,802	0.01	No	–	–	0.202	0.173	0.230
Processes	CC	1,836	0.01	No	0.030	Yes	0.050	0.027	0.078
	CR	1,848	0.06	Yes	0.057	Yes	0.087	0.057	0.120
	DR	1,847	0.06	Yes	0.099	Yes	0.083	0.054	0.116
	ER	1,830	0.06	No	0.017	No	0.126	0.091	0.163
	HR	1,847	0.01	No	0.004	No	0.035	0.016	0.059
	HT	1,842	0.01	No	–	–	0.277	0.234	0.319
	IS	1,831	0.06	No	–	–	0.123	0.088	0.160
	RE	1,845	0.06	No	–	–	0.235	0.193	0.276
	RM	1,835	0.01	No	0.010	No	0.056	0.031	0.084
	RG	1,838	0.14	No	–	–	0.059	0.034	0.088
	SL	1,845	0.14	Yes	0.101	No	0.166	0.127	0.205
	SR	1,845	0.14	No	0.183	No	0.101	0.069	0.135
SS	1,848	0.06	No	–	–	0.135	0.099	0.172	
SO	1,845	0.01	Yes	0.000	No	0.013	0.002	0.030	
HC	1,802	0.01	No	–	–	0.038	0.018	0.063	

ω^2_{pred} predicted effect size, ω^2_{obs} observed effect size, *L-CI* lower confidence interval, *U-CI* upper confidence interval, *Sunsc.* sunscreen confidence, *Sun Av.* sun avoidance confidence, *CC* counter conditioning, *CR* consciousness raising, *DR* dramatic relief, *ER* environmental re-evaluation, *HR* helping relationships, *HT* health responsibility, *IS* interpersonal systems control, *RE* reducing exposure, *RM* reinforcement management, *RG* regret, *SL* self-liberation, *SR* self-re-evaluation, *SS* sunscreen use, *SO* social liberation, *HC* health care provider

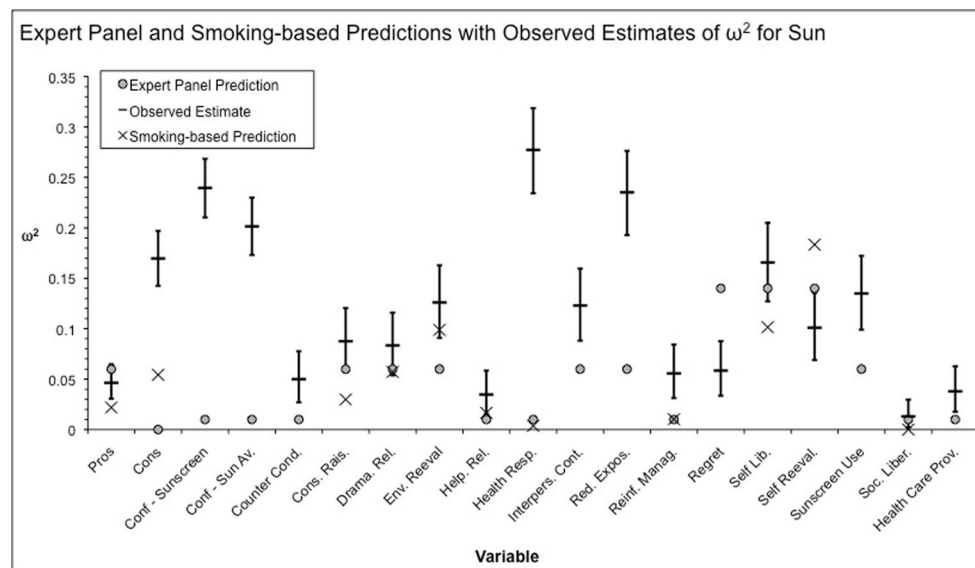


Fig. 3 Comparison of expert panel and smoking-based predictions with observed ω^2 effect size estimates for sun protection constructs surrounded by 99 % confidence intervals. *Conf-Sunscreen* sunscreen confidence, *Conf-Sun Av.* sun avoidance confidence, *CC* counter conditioning, *CR* consciousness raising, *DR* dramatic relief, *ER*

environmental re-evaluation, *HR* helping relationships, *HT* health responsibility, *IS* interpersonal systems control, *RE* reducing exposure, *RM* reinforcement management, *RG* regret, *SL* self-liberation, *SR* self-re-evaluation, *SS* sunscreen use, *SO* social liberation, *HC* health care provider

sizes, negating a need for consideration of another category of “extra large” effects. Revision using an “extra large” category would be able to better describe four of remaining missed predictions (i.e., sunscreen use confidence, sun avoidance confidence, HR, and RE).

Discussion

One of the main goals of the TTQP approach is promote movement away from NHST by requiring researchers to make specific effect size predictions based on theory and empirical evidence. The approach also serves to foster the use of confidence interval estimation and replication. Major findings from this study demonstrate that the TTQP approach works best when predictions are derived from previously existing empirical data, such as the case with smoking results. In addition, the study found that the expert panel did very poorly in generating precise expected effect size values. These findings suggest that future studies utilizing the TTQP approach should do their best to generate predictions based on previous research and thoroughly consider how sample data serves to update expectations.

Study 1 represents the third replication for smoking behavior variables in the TTM and demonstrates that with multiple iterations of theoretical predictions, expected values begin to converge on the population effect size. This is important, as it helps build and synthesize empirical support for a strong theory. Results from studies 2 and 3 demonstrate the difficulty in generating accurate predictions of effect sizes. Both the smoking-based and the expert panel approaches produced

high numbers of missed predictions. In study 2, about half of the smoking-based predictions directly transferred to diet. Slightly better results were obtained for diet using smoking-based predictions compared to expert panel predictions. Results from study 3 indicate that effect sizes for sun protection were strikingly different than those observed for smoking. For sun protection overall, both the expert panel and the smoking-based effect size predictions were much lower than observed effects and resulted in a large number of misses.

The successful confirmation of predictions from study 1 lends substantial support for the cross-sectional TTM theory predictions in smoking behavior. However, this is the first application of the TTQP approach on TTM measures of sun protection and diet in studies 2 and 3. The inadequate fit of smoking-based predictions suggests that diet and sun protection constructs have substantially different effects across the first three stage of change compared to smoking. Observed estimates from studies 2 and 3 should be used as starting values of predictions in an independent sample to provide a second iteration of the approach using empirical data specific to each behavior rather than expert panel or smoking-based estimates. Findings from all three studies demonstrate the iterative nature of this approach and the need for replication using independent samples.

The TTQP approach highlights the degree to which measured constructs differentially affect health behavior, allowing comparison of theory across various behavioral areas. This information is valuable for future studies that apply the TTM, or other competing theories, to target specific behaviors. For example, findings from studies 1, 2, and 3 reveal that the

construct of self-efficacy showed the highest effect size for sun protection, suggesting that people in the first three stages of change vary more in their confidence that they can protect their skin than in their temptation to smoke or eat high fat foods.

A few limitations to the TTQP approach in this sample should be noted. All of the studies conducted in this paper were cross-sectional and used baseline measures, representing only a snapshot of the effects specific to each construct. While baseline information is valuable, the TTQP can also be extended as a longitudinal approach to examine the effects of constructs across time. This approach has been applied to smoking behavior [16] at different intervention time-points, but needs to be replicated and applied to new behaviors. Given the lack of diversity in this sample and effort to examine effect sizes across population groups would also be beneficial. Finally, as demonstrated by the great variability in effect sizes across the three studies, this approach requires fairly specific predictions regarding the constructs, measures, sample, and behaviors. Such specificity should advance science by creating a more clearly empirical foundation for the accumulation of knowledge. Theory and context should always guide predictions and close examination of misses should always be conducted.

The methodology presented thus far represents an alternative to NHST via the promotion of a quantitative oriented approach. By using effect sizes and CIs to test quantitative predictions, researchers are able to bypass significance tests and acquire more information about effects (e.g., magnitude and CI) than from a dichotomous accept–reject framework based on p values. Effect sizes used as the basis for predictions should be derived from previous research. Meta-analyses, in particular, would serve as an especially strong basis for deriving theoretical-based effect size predictions [37]. The TTQP approach described here shifts away from the traditional way of thinking by emphasizing the effect size of a construct, making it an especially useful tool for intervention development using an empirically supported theoretical framework. This approach also allows for straightforward comparison across studies and across theories by utilizing a common effect size metric, thus facilitating meta-analyses and integration into novel behavioral interventions.

Acknowledgments Funding for this study was partially provided by National Cancer Institute (NCI) Grant CA11919, National Institute on Drug Abuse (NIDA) Grant DA02112, and National Center for Research Resources (NCRR) Grant G20RR030883. NCI, NIDA, and NCRR played no role in the study design, collection, analysis or interpretation of the data, writing of the manuscript, or the decision to submit the paper for publication.

Ethical Standards All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000. Informed consent was obtained from all participants for being included in the study.

Conflict of Interest All authors declare no conflicts of interest.

References

1. Kline RB. Beyond significance testing: reforming data analysis methods in behavioral research. Washington DC: American Psychological Association; 2004.
2. Rossi JS. Statistical power analysis. In: Schinka J, Velicer WF, editors. Handbook of psychology. Volume 2: research methods in psychology. Wiley; 2013.
3. Gigerenzer G. Mindless statistics. *J Socio-Econ.* 2004;33(5):587–606.
4. Steiger JH. Beyond the F test: effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychol Methods.* 2004;9(2):164.
5. Cumming G. Understanding the new statistics: effect sizes, confidence intervals, and meta-analysis. New York: Routledge Academic; 2012.
6. Velicer WF et al. Theory testing using quantitative predictions of effect size. *Appl Psychol.* 2008;57(4):589–608.
7. Rossi JS. Contrast effect size analysis and theory testing: an example using the transtheoretical model. *Ann Behav Med.* 2001;23(S1):s156.
8. Rossi JS. Comparison of the use of significance testing and effect sizes in theory-based health promotion research. In: 43rd annual meeting of the Society of Multivariate Experimental Psychology. Keystone; 2003.
9. Velicer WF et al. Using the transtheoretical model for population-based approaches to health promotion and disease prevention. *Homeostasis Health Dis.* 2000;40(5):174–95.
10. Prochaska JO. Stages and processes of self-change of smoking: toward an integrative model of change. *J Consult Clin Psychol.* 1983;51(3):390–5.
11. Prochaska JO, DiClemente CC. The transtheoretical approach: crossing the traditional boundaries of therapy. Melbourne: Krieger Publishing Company; 1984.
12. Prochaska JO. Transtheoretical model of behavior change. In: Encyclopedia of behavioral medicine. Springer; 2013. p. 1997–2000.
13. Prochaska JO et al. Stages of change and decisional balance for 12 problem behaviors. *Health Psychol.* 1994;13(1):39.
14. Hall KL, Rossi JS. Meta-analytic examination of the strong and weak principles across 48 health behaviors. *Prev Med.* 2008;46(3):266–74.
15. Velicer WF et al. Testing 40 predictions from the transtheoretical model. *Addict Behav.* 1999;24(4):455–69.
16. Velicer WF et al. Testing 40 predictions from the transtheoretical model again, with confidence. *Multivar Behav Res.* 2013;48(2): 220–40.
17. Fava JL, Velicer WF, Prochaska JO. Applying the transtheoretical model to a representative sample of smokers. *Addict Behav.* 1995;20(2):189–203.
18. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale: Erlbaum; 1988.
19. DePue JD et al. Cancer prevention in primary care: predictors of patient counseling across four risk behaviors over 24 months. *Prev Med.* 2008;46(3):252–9.
20. Prochaska JO et al. Stage-based expert systems to guide a population of primary care patients to quit smoking, eat healthier, prevent skin cancer, and receive regular mammograms. *Prev Med.* 2005;41(2):406–16.
21. Greene GW et al. Baseline transtheoretical and dietary behavioral predictors of dietary fat moderation over 12 and 24 months. *Eat Behav.* 2013;14(3):255–62.
22. Greene GW et al. Stages of change for reducing dietary fat to 30% of energy or less. *J Am Diet Assoc.* 1994;94(10):1105–12.

23. Rossi J et al. Behavior change for reducing sun and ultraviolet light exposure: implications for interventions. *Dermatol Clin*. 1995;13: 613–22.
24. Weinstock MA et al. Randomized controlled community trial of the efficacy of a multicomponent stage-matched intervention to increase sun protection among beachgoers. *Prev Med*. 2002;35(6): 584–92.
25. DiClemente CC et al. The process of smoking cessation: an analysis of precontemplation, contemplation, and preparation stages of change. *J Consult Clin Psychol*. 1991;59(2):295.
26. Greene GW et al. Dietary applications of the stages of change model. *J Am Diet Assoc*. 1999;99(6):673–8.
27. Rossi J. et al. Effectiveness of stage-matched interventions for skin cancer prevention: a randomized clinical trial of high-risk beach bathers. *Ann Behav Med*. 1997;19(S194).
28. Velicer WF et al. Decisional balance measure for assessing and predicting smoking status. *J Pers Soc Psychol*. 1985;48(5):1279.
29. Diclemente CC. Self-efficacy and the addictive behaviors. *J Soc Clin Psychol*. 1986;4:302–15.
30. Velicer WF et al. Relapse situations and self-efficacy: an integrative model. *Addict Behav*. 1990;15(3):271–83.
31. Rossi SR et al. Validation of decisional balance and situational temptations measures for dietary fat reduction in a large school-based population of adolescents. *Eat Behav*. 2001;2(1):1–18.
32. Prochaska JO et al. Measuring processes of change: applications to the cessation of smoking. *J Consult Clin Psychol*. 1988;56(4):520.
33. Santiago-Rivas M, Velicer WF, Redding C. Mediation analysis of decisional balance, sun avoidance and sunscreen use in the precontemplation and preparation stages for sun protection. *Psychology & Health*. 2015;24:1–17.
34. Sillice MA et al. (Abstract) Psychometric assessment of the processes of change scale for sun protection. *Ann Behav Med*. 2014;47(S1):s247.
35. Yusufov M, et al. Transtheoretical model constructs' longitudinal prediction of sun protection over 24 months. *Int J Behav Med*. 2015; in press.
36. Kromrey JD, Bell BA. ES_ANOVA: a SAS® macro for computing point and interval estimates of effect sizes associated with analysis of variance models. In: SouthEast SAS Users Group 2010 Proceedings. Savannah; 2010.
37. Rossi JS. A case study in the failure of psychology as a cumulative science: the spontaneous recovery of verbal learning. In: Harlow LL, Mulaik SA, Steiger JH, editors. What if there were no significance tests? Mahwah: Lawrence Erlbaum; 1997. p. 175–97.