Journal of Data Science **4**(2006), 291-306

An Evaluation of Multiple Behavioral Risk Factors for Cancer in a Working Class, Multi-Ethnic Population

Melody S. Goodman[1], Yi Li[1,2], Gary G. Bennett[1,2],
Anne M. Stoddard[3], and Karen M. Emmons[1,2]
[1]*Harvard School of Public Health* [2]*Dana Farber Cancer Institute and* [3]*New England Research Institute*

Abstract: Behavioral risk factors for cancer tend to cluster within individuals, which can compound risk beyond that associated with the individual risk factors alone. There has been increasing attention paid to the prevalence of multiple risk factors (MRF) for cancer, and to the importance of designing interventions that help individuals reduce their risks across multiple behaviors simultaneously. The purpose of this paper is to develop methodology to identify an optimal linear combination of multiple risk factors (score function) which would facilitate evaluation of cancer interventions.

Key words: Community based research, conditional logistic regression, multiple risk factors, random effects.

1. Introduction

Despite the considerable biomedical advances of the last half-century, facilitating improvement in lifestyle behaviors remains the most efficacious population-level strategy for reducing cancer risk. Estimates vary, but suggest that over fifty percent of new cancer cases and up to one-third of cancer mortality could be prevented though improvements in health behavior practices (American Cancer Society, 2004; Doll and Peto, 1981). A 19 percent decline in the rate at which new cancer cases occur, and a 29 percent decline in the rate of cancer deaths, could potentially be achieved by 2015, if prevention efforts were heightened and behavior change sustained. This would translate to the prevention of approximately 100,000 cancer cases and 60,000 cancer deaths each year, by the year 2015 (National Cancer Policy Board and Institute of Medicine, 2003).

There is ample epidemiological evidence for the consideration of red meat consumption, physical activity, and folic acid intake in cancer prevention efforts. Regular physical activity lowers the risk of cancers of the colon, breast, and possibly prostate (Colditz, Cannuscio, and Frazier, 1997; Friedenreich and Rohan, 1995). An additional 30 percent of cancer deaths can be attributed to adult

diet (Anonymous, 1996); higher intake of red meat has been associated with increased risk of colon (Sandhu, White and McPherson, 2001) and prostate cancers (Michaud, Augustsson, Rimm, Stampfer, Willett, and Giovannucci 2001). Associated with both physical inactivity and diet is obesity, which may account for between 25-30 percent of cancers of the colon, breast (postmenopausal), endometrium, kidney, and esophagus (Vainio and Bianchini, 2002). Folic acid is protective against colon cancer (Giovannucci, Stampfer, Colditz, Hunter, Fuchs, Rosner, Speizer, and Willett, 1998); long-term multi-vitamin use, in particular has been found to reduce risk for colon cancer, likely because of its folic acid content (Giovannucci *et al.*, 1998).

The risk for many diseases, including colon cancer, is associated with multiple behavioral risk factors (MRF); these behaviors are highly interrelated and tend to cluster within individuals. For example, those who eat high-fat diets are also more likely to be sedentary, suggesting that the behaviors may be mutually reinforcing (see e.g., Emmons, Marcus, Linnan, Rossi, and Abrams, 1999). Change in one behavioral risk factor thus may serve as a stimulus or gateway for change in the other health behaviors (see e.g., Emmons *et al.*, 1999), and there are overarching behavioral principles and intervention frameworks that guide behavior change efforts across risk factors. Consequently, to facilitate population-level reductions in cancer risk, it may be inefficient to target discrete behavioral risk factors, when similar principles might be applied simultaneously to multiple behaviors (Institute of Medicine, 2000).

The literature provides little consensus as to the most appropriate analytic strategy for evaluating the efficacy of MRF interventions; most studies have analyzed the various outcomes independently or by creating a simplistic sum (e.g., 1 RF + 1 RF = 2RFs) (see e.g., Prochaska and Sallis 2004; Campbell, James, Hudson, Carr, Jackson, Oakes, Demissie, Farrell, and Tessaro, 2004). This could be problematic, because the use of separate analytic strategies may result in improper inferences regarding the effect of an MRF intervention because of correlation among the factors. Such strategies may overlook the clustering effect brought about by the agglomeration of multiple behavioral risk factors and have been criticized as being too simplistic. The purpose of this paper is to develop a methodology to identify an optimal linear combination of multiple behavioral risk factors (MRF score function) for cancer that would best facilitate evaluation of an MRF cancer intervention.

2. Methods

2.1 Study design

The data analyzed in this paper are from the Harvard Cancer Prevention Pro-

gram Project (HCPPP) Healthy Directions, which is composed of two randomized controlled trials, one in health centers (HC) (Emmons, Stoddard, Gutheil, Suarez, Lobb, and Fletcher 2003), and another in small businesses (SB) (Hunt, Stoddard, Barbeau, Wallace, and Sorensen 2003). The overarching goal of the HCPPP was to create a new generation of cancer prevention interventions that would be effective among working class, multi-ethnic populations. Together, the two arms of the trial were successful in enrolling a sub-population of the multi-ethnic working class population in eastern Massachusetts. The study aims and sampling strategies are published in greater detail elsewhere (Emmons *et al.*, 2003; Hunt *et al.*, 2003).

2.2 Health centers

Healthy Directions-HC (Emmons *et al.*, 2003) was a randomized controlled trial conducted in collaboration with a large health care delivery system, comprised of 14 multi-specialty medical group practices that serve over 270,000 patients. Ten of the fourteen health centers were invited to participate in this study, and all agreed. Health center served as the unit of randomization and intervention. Briefly, patients who resided in low income, multi-ethnic neighborhoods (defined using census block-groups that were predominantly working class, impoverished, or with low levels of education) were identified and approached for participation through their health center. Individuals identified through geocoding to be residents in the target neighborhoods were deemed eligible if they met the following criteria: (1) being 18-75 years old, (2) having a well-care or follow-up visit scheduled with a participating provider, (3) being able to speak and read either English or Spanish, (4) not having cancer at the time of enrollment, (5) not being employed by the participating health centers, (6) not being employed by a worksite participating in the companion small business study, and (7) providing consent to participate in the randomized study. All providers practicing in the Internal Medicine Departments of the health centers were approached for permission to recruit from among their patient pools. Provider participation averaged 83% across sites (range 50%-100%; 97 clinicians). Patients scheduled for appointments with the participating providers and in the eligible age range were identified through the automated central appointment system. Study staff attempted to recruit 8,963 potentially eligible candidates; 2,547 (28%) individuals were unreachable. Among the 6,414 potential participants reached, 867 (14%) were ineligible, 3,330 (52%) refused, and 2,219 (35%) were enrolled. Assuming that 14% of those not reached were also ineligible, the response rate is 29% of those assumed eligible. The cohort recruited at baseline was contacted by telephone after the intervention period to complete a follow-up survey. Of the 2,219 who completed the baseline survey (n=1088 intervention condition; n=1131

control condition), 1,954 (88%) completed the follow-up survey. The follow-up response rate was equivalent across conditions.

2.3 Small business

The Healthy Directions-SB study (Hunt *et al.*, 2003) was a randomized controlled trial in which the worksite was the unit of randomization and intervention. Worksites were identified using the Dun and Bradstreet database to locate small businesses with Standard Industrial Classification (SIC) codes 20-39 (manufacturing industries) and employing between 50-150 employees. Additional inclusion criteria included: (1) employing a multi-ethnic population (defined as 25% of workers being first- or second-generation immigrants or people of color), (2) having a turnover rate of less than 20% in the previous year, (3) being autonomous in decision-making power to participate in a study, and (4) agreeing to be randomly assigned to the intervention condition. One hundred thirty-three (133) companies met the eligibility criteria, and of these, 26 agreed to participate (Barbeau, Wallace, Lederman, Lightman, Stoddard and Sorensen, 2004).

Data were collected using interviewer-administered surveys among individuals who were permanent employees and worked 20 hours or more per week. On site interviews were administered on company time in the language (either English, Spanish, Portuguese, or Vietnamese) preferred by respondents. Two cross-sectional samples were collected, one at baseline in which 1,740 participants from 26 worksites completed the survey (response rate 84%). The second sample was collected at follow-up, comprising 1,408 participants in 24 worksites (during the course of the intervention two worksites dropped out, one intervention and one control) with a response rate of 77%. 974 participants (518 in control worksites and 456 in intervention worksites) completed both the baseline and follow-up surveys forming the embedded cohort used in this analysis.

2.4 Data and analysis

The goals of the intervention were to: (1) increase fruit and vegetable intake, (2) decrease red meat consumption, (3) increase physical activity levels, and (4) increase daily multivitamin usage. The following variables assess the individual risk factors measured on a continuous scale: number of servings of fruit and vegetables per day, number of servings of red meat consumed per week (RM), and hours of moderate or vigorous physical activity per week (PA). The fourth measure is a binary variable indicating use of a multi-vitamin on 6 or 7 days per week (MV). In order to keep all variables on an equivalent time scale, we created a new variable for fruit and vegetable consumption that calculated the amount of fruits and vegetables consumed in one week (FV) by multiplying the current

measure of fruit and vegetable intake by seven. The continuous variables (FV, RM, PA) were standardized using the formula in Equation 2.1;

$$STV = \frac{V - P_{05}}{P_{95} - P_{05}}, \tag{2.1}$$

where V are the original values for the continuous variables (FV, RM, PA), P_{05} and P_{95} are the fifth and ninety-fifth percentile values respectively for a given variable and STV are the new standardized variables (STFV, STRM, and STPA respectively). Standardization was implemented for consistency (to make a one unit change in one variable similar to a one unit change in another) and interpretability. The 5th and 95th percentiles were used to minimize the influence of outliers.

For the purposes of identifying an optimal linear combination that would show an intervention effect we restricted our sample to only those participants who received the intervention, responded to both the baseline and follow-up surveys, and have complete data for the four risk behaviors. As opposed to the usual situation of observing how the covariate vector or a linear combination of the covariate vector will change because of treatment, the idea here is to determine how the covariate vector or the linear combination will predict the intervention status. This is similar in spirit to a matched case-control analysis.

A popular method for the analysis of longitudinal data with a dichotomous outcome is a mixed effects logistic regression model. A mixed effects logistic regression model with a logit link will have the form:

$$\log\left[\frac{pr(Y_{ij} = 1)}{1 - pr(Y_{ij} = 1)}\right] = a_i + \beta' X_{ij}. \tag{2.2}$$

Here Y_{ij}, $i = 1, ..., n$, $j = 1, 2$, denotes the indicator of intervention time (i.e. pre-intervention $Y_{i1} = 0$ and post-intervention $Y_{i2} = 1$), X_{ij} is the covariate vector, and a_i is a random cluster effect. The subscript i is an indicator for individual and the subscript j is an indicator for time. Each individual subject i is a cluster of two sets of observations, pre-intervention and post-intervention. The random effect variable a_i can be thought of as measuring an individual's demographic characteristics (i.e., age, gender, race). In our analysis, we want to control for an individual's specific demographic characteristics, therefore, we treat the random effect variable a_i as a nuisance parameter and condition it out of the model. We can condition them out by using the conditional likelihood based on the fact that $Y_{i1} + Y_{i2} = 1$. We are left with a conditional logistic regression model. These types of models are often used to analyze matched case-control studies, where the outcome of interest is whether a participant is a case or control.

In this framework we intend to model

$$\text{logit}(Pr(Y_{ij} = 1|a_i)) = \beta' X_{ij}, \tag{2.3}$$

where an optimal linear combination, or the best score, will be $\hat{\beta}' X$.

We set up our data as if it came from a 1:1 matched case-control study; each individual is a cluster of two observations, one "case" and one "control". One observation is pre-intervention ("control"/baseline) and the second observation is post-intervention ("case"/follow-up). At each time point (pre and post-intervention) each participant has a vector (containing STFV, STRM, STPA, and MV) of covariates.

For matched case-control studies with one case per matched set, the likelihood function for the conditional logistic regression reduces to the partial likelihood of the Cox model for the continuous time scale (Hosmer and Lemeshow 1998). We created dummy survival times so that all cases have the same event time and the corresponding controls are censored at a later time. We used PROC PHREG in SAS/STAT™ software to fit the conditional logistic regression model by forming a stratum for each matched set (individual id number). This allowed us to obtain estimates for $\hat{\beta}$.

3. Results

Using the combined Health Center and Small Business data from the Healthy Directions baseline and follow-up surveys on the 1,209 study participants that received the intervention, we found an optimal score function for the four risk factors:

$$score = 1.05 * STFV + 1.70 * MV + 0.25 * STPA - 1.35 * STRM. \tag{3.1}$$

The score is a summary measure of the health behaviors of a participant based on these four factors. From this score, we can see that increasing the number of fruits and vegetables consumed per week, taking a multivitamin six or more days a week, increasing the amount of physical activity done in a week, and/or decreasing the amount of red meat consumed in a week will increase the score for a participant which in turn means an overall improvement in health behaviors. The dynamics of the score are consistent with the goals of the intervention. A participant can increase their health behavior score by changing one risk factor, or combinations of the four risk factors in a manner consistent with the goals of the intervention.

We believe that these factors not only have individual effects, but that some factors may also have compounding effects. This belief is based on previous evidence of the interrelationships seen in modifying behavioral risk factors (see e.g., Emmons *et al.*, 2004; Butterfield *et al.*, 2004). Therefore, we looked for

significant interactions between the four variables. Table 1 shows the analysis of maximum likelihood estimates for our final model. In our final score function (see Equation 3.2), we multiply the effects (parameter estimates) by 100 to increase the range of the scores as well as to simplify interpretation.

Table 1: Analysis of Maximum Likelihood Estimates

Standardized Variable	Parameter Estimate	Standard Error	P-Value
STFV	0.576	0.303	0.0570
MV	2.008	0.2078	<.0001
STPA	0.232	0.193	0.2294
STRM	-1.515	0.343	<.0001
STFV*STRM	1.229	0.565	0.0296
MV*STRM	-0.707	0.343	0.0392

Table 2: Examples of changes in risk factor measures and resulting MRF score

	FV	MV	PA	RM	MRF Score
Baseline values	20	0	4	5	−30.23
Case 1: Optimal values at final					
Final values	35	1	10	1	247.46
change	+15	+1	+6	−4	+277.69
Case 2: Improves only FV					
Final values	35	0	4	5	17.39
change	+15	0	0	0	+47.62
Case 3: Improves only MV					
Final values	20	1	4	5	135.22
change	0	+1	0	0	+165.45
Case 4: Improves only PA					
Final values	20	0	10	5	−19.09
change	0	0	+6	0	+11.14
Case 5: Improves Only RM					
Final values	20	0	4	1	14.64
change	0	0	0	−4	+44.87
Case 6: Improves FV and RM					
Final values	35	0	4	1	72.82
change	+15	0	0	−4	+103.05

$$\begin{aligned} \text{score} &= 57.6 * STFV + 200.8 * MV + 23.2 * STPA - 151.5 * STRM \\ &\quad + 122.9 * STFV * STRM - 70.7 * MV * STRM \end{aligned} \tag{3.2}$$

There was a significant interaction between the amount of fruits and vegetables consumed per week and the amount of red meat consumed per week, sugessting that changing both behaviors simultaneously is better than changing either behavior alone, but the effect of changing both behaviors is not equal to the sum of the individual changes on the MRF score. There was also a significant interaction between multivitamin usage more than six times a week and the amount of red meat consumed per week, suggesting that changing either behavior alone is good, but changing both behaviors simultaneously will result in an even larger increase on the MRF score.

Table 3: Score changes with one, two, and three variable changes

Variables Changed	Score Change
FV	47.62
MV	165.45
PA	11.14
RM	44.87
FV + MV	213.07
FV + PA	58.76
FV + RM	72.82
MV + PA	176.59
MV + RM	238.60
PA + RM	56.00
FV + MV + PA	224.21
FV + MV + RM	266.55
FV + PA + RM	83.96
MV + PA + RM	249.73

Table 2 displays a few examples of how a change in an individual risk factor from the baseline case to the optimal case will change the score. If we consider the first row of Table 2 to be a baseline value in which a participant consumes 20 servings of fruits and vegetables per week, does not take a multivitamin six or more days a week, has four hours of physical activity per week, and consumes five servings of red meat per week (the average values for study participants at baseline, meeting only the recommend level of physical activity), the standardized values would be 0.32,0,0.32, and 0.5 respectively. Therefore the score for a participant at baseline would be

$$\begin{aligned} score &= 57.6 * 0.32 + 200.8 * 0 + 23.3 * 0.32 - 151.5 * 0.5 \\ &\quad + 122.9 * 0.32 * 0.5 - 70.7 * 0 * 0.5 = -30.2. \end{aligned} \tag{3.3}$$

We can consider an arbitrary optimal case as a participant who consumes 35 servings of fruits and vegetables per week (or five a day), takes a multivitamin 6 or more days a week, engages in 10 hours of physical activity per week, and eats one serving of red meat per week (meeting and/or exceeding all of the recommended levels). Table 3 shows the effects of these changes on the score from the baseline case to the optimal case for each variable alone and the effects of combinations of two and three variables. Figure 1 compares our final model (MRF, Equation 3.2) with a main effects model (a model without interactions) showing that the main effects model can both overestimate and underestimate scores predicted from the MRF model due to the absence of the two significant interactions.

Although we used only those participants that received the intervention to develop the score, the score is generalizable to the entire study population. It was created, and is most useful for, the purpose of comparing the participants that received the intervention to those that received usual care, because it provides a summary measure of the health behaviors of a participant on all intervention risk factors pre and post-intervention. There were 1,297 participants that received usual care and took both the baseline and follow-up surveys. These participants can be considered controls for the effect of the intervention. Figure 2 shows box plots of MRF score comparing baseline and follow-up for participants that received the intervention compared to those that received usual care. In the intervention group, the mean score at baseline was 48.1, while the mean score at follow up was 104.3. In the usual care group the mean score at baseline was 40.4, and the mean score at follow-up was 53.2. The mean change in score for the usual care group was 12.8, while the mean change in score for the intervention group was 56.2. There was a statistically significant difference in the mean change in score from baseline to follow-up when comparing the usual care group to the intervention group (p <0.001). The intervention group showed greater improvements in MRF score at follow-up proving the intervention successful.

4. Discussion

Increasing attention has been paid to multiple risk factor interventions, across a range of disease outcomes, both because adverse behavioral risk factors tend to cluster within individuals and because of recognition of the utility of facilitating change across multiple risk behaviors. However, most MRF studies to date have used individual risk factor methods to analyze intervention effects (see

e.g., Prochaska and Sallis, 2004; Campbell *et al.*, 2004). As shown in Figure 1, the main effects model both over-estimates (e.g., FV & PA & RM) and under-estimates (e.g., MV & PA & RM) the scores predicted from the MRF model, depending on the combination of variables and the degree of change for a given participant in the intervention. Thus, such analytic models may compromise determinations of the efficacy of a MRF intervention. We were successful in modeling a linear combination of behavioral risk factors including interactions between risk factors, an effort that represents an advance over the existing methods for analyzing MRF intervention efficacy.

To illustrate, note that in our final model there are two interaction terms. One between the amount of fruits and vegetables consumed per week and the amount of red meat consumed per week, and another between multivitamin usage more than six times a week and the amount of red meat consumed per week. Looking at Table 3, we can see that with all the other variables held constant, a change in fruit and vegetable consumption alone from 20 to 35 servings per week will increase the score by 47.62, and a decrease in red meat alone from 5 to 1 servings per week will increase the score by 44.87. However, because of the interaction term, if both variables are changed by the amounts indicated the score would increase by 72.82, which because of the interaction is a smaller than 92.49, the sum of the individual changes. Similarly, if a participant begins to take a multivitamin daily the score will increase by 165.45, and if they decrease red meat from 5 to 1 serving per week the score will increase by 44.87. However, if a participant begins to take a multivitamin daily and decreases red meat consumption by 4 servings per week the score will increase by 238.60, a larger increase than 210.32 that you would get by adding 165.45 from taking a multivitamin daily and 44.87 by decreasing red meat consumption. Cluster effects are not captured by main effects models and are an advantage of this method.

There are some limitations to the method proposed here, namely that the score function depends on the efficacy of the intervention to determine variable weighting. For example, if the intervention was most effective at increasing multivitamin use, the weight (coefficient) for the multivitamin use variable would be largest in magnitude, whereas if the intervention was least effective in changing the participants' physical activity patterns, the weight (coefficient) for the physical activity variable would be the smallest in magnitude. In some cases then, the weights may be a proxy for the amount of participant effort necessary to change the health behavior. For example, in this study we saw that multivitamin usage had the largest weight and thus the most influence on the score.

There are at least two potential explanations for this finding. First, the promotion of multivitamin usage may require less participant burden when compared to the other health behaviors (e.g., physical activity). Thus, it may be easier for

participants to modify their multivitamin use; this supposition appears to be supported by the finding of an almost thirty percent increase from baseline to follow-up of the number of participants taking a multivitamin daily. However, it is important not to undermine the significance of a change in multivitamin usage which is strongly related to the prevention of disease outcomes. Sustained use of multivitamins containing folic acid have been associated with the reduction in risks for numerous conditions including colorectal cancer and cardiovascular disease (Ggiovannucci *et al.*, 2002; Fairfield and Fletcher, 2002). Physical activity on the other hand, is among the most challenging health behaviors addressed in the study to intervene upon. In this population, 66 percent of the participants were getting the recommended level of physical activity at baseline, and 69 percent at follow-up. Of those participants that were not at or above the target level of physical activity at baseline, almost 9.5 percent were at or above the target level at follow-up. Another factor to consider is that multivitamin usage was treated as a binary variable in our models. That is, many potential changes are captured in the categorization of either taking a multivitamin 6 or more times a week or not doing so. Relative to increasing one serving of fruits and vegetables a week, decreasing one serving of red meat in a week, or increasing an hour of physical activity a week, this is a substantial change.

Although the purpose of our method was to develop a health behavior score (composite variable), there are some limitations to using this type of variable. The purpose of such a variable is to allow for easy comparisons of the four factors with one number. When there are changes in the score, however, a composite variable does not provide any insight into which individual risk factor(s) have contributed to the change.

Another potential limitation of applying this method to the HCPPP data is the merging of the two cohorts, small businesses and health centers. Our method develops a score function that is independent of the population but not independent of the intervention. By combining the two cohorts, we have made the assumption that the interventions given to these two populations are the same. In reality, although the two interventions were quite similar, they were not identical. We decided, however, to combine the two cohorts in order to increase power, and to create a universal score that could be applied to both cohorts. This not only allowed us to make comparisons within a cohort, but between cohorts. Taking these limitations into account, our methodology remains preferable compared to existing techniques that do not accord weights to the risk factors or adjust for cluster effects.

In summary, we have developed a score that effectively integrates multiple behavioral cancer risk factors into one measure, irrespective of individual demographic factors. We believe that the methods are generalizable to other working

class multi-ethnic populations, and future research should be done to evaluate the effectiveness of these methods in other groups. The primary strength of the methodology used to develop the score is that it can be easily implemented to develop scores for other populations, for other combinations of behavioral risk factors, or for other disease outcomes (e.g., cardiovascular disease). Given the increasing attention being paid to the development of MRF interventions, we believe the described method to be the preferred means of analysis in comparison to previously used strategies. Ultimately, we believe that analytic focus on examining clusters of behavioral risk factors will enhance the design of multiple risk factor intervention approaches.

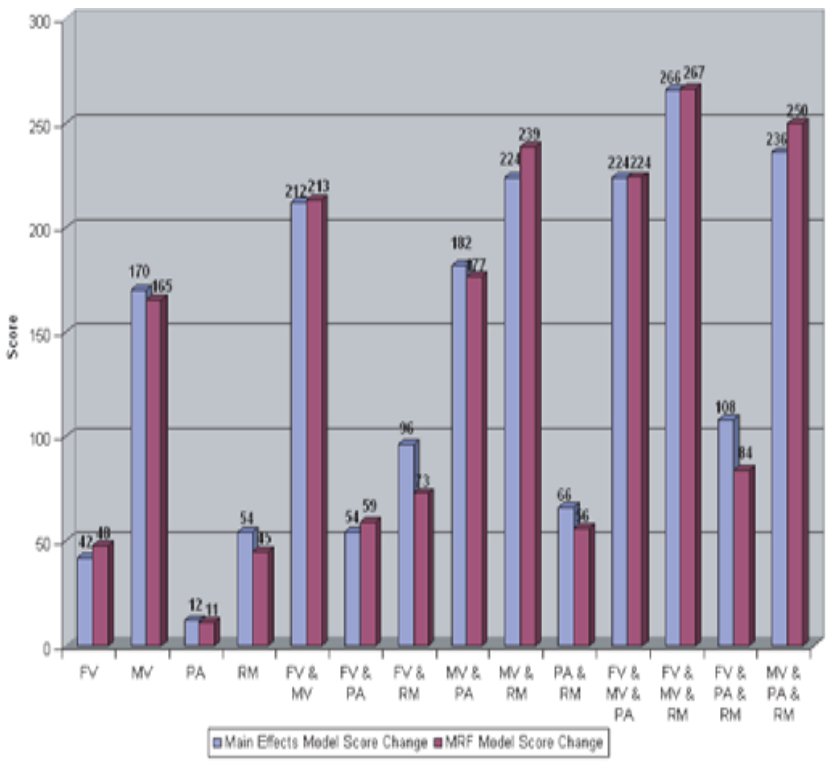

Figure 1: Comparison of Model Score Changes

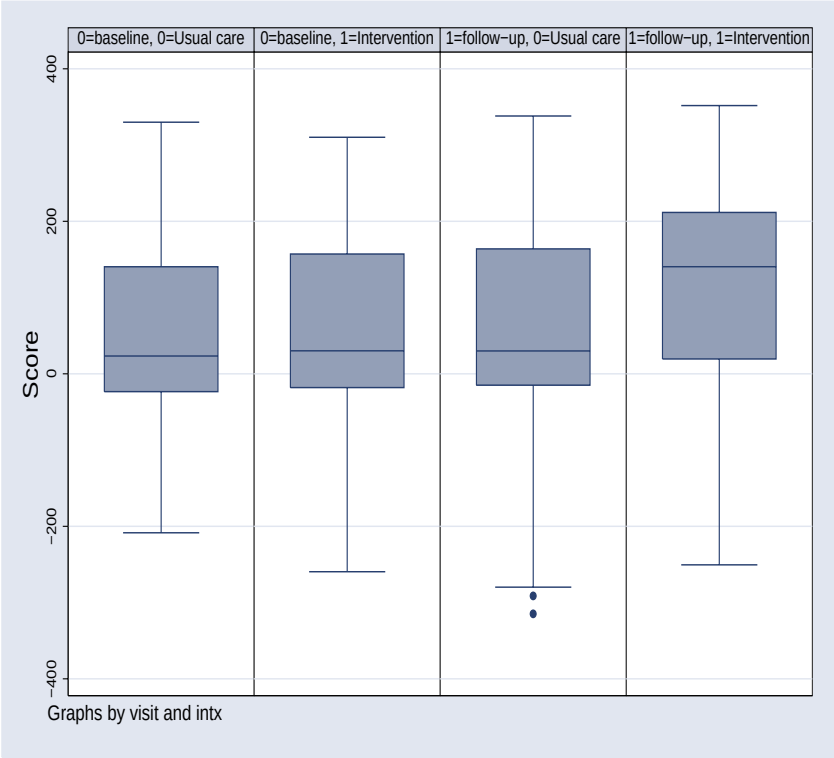

Figure 2: Comparison of scores for participants by visit and intervention status

Acknowledgements

The research of Dr. Melody Goodman was supported by National Institute of Child Health and Human Development grant 5 F31 HD043695. The research of Dr. Li was supported by National Institute of Health grant R01CA95747. The research of Dr. Bennett, Dr. Stoddard, and Dr. Emmons was supported by grant 5 P01 CA75308 from the National Institutes of Health, and support to the Dana-Farber Cancer Institute by Liberty Mutual, National Grid, and the Patterson Fellowship Fund.

The statistical output for this paper was generated using SAS/STAT software, Version 8 of the SAS System for Windows. Copyright@ 1999-2001 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

References

American Cancer Society (2004). Cancer Facts and Figures. Technical Report, American Cancer Society.

Anonymous (1996). Harvard Report on Cancer prevention Volume: 1 Causes of Human Cancer, *Cancer Causes and Control* **7**, S3-S9.

Barbeau, E. M., Wallace, L., Lederman, R., Lightman, N., Stoddard, A. and Sorensen, G. (2004). Recruiting small manufacturing worksites that employ multi-ethnic, low-wage workers to a cancer prevention research trial. *Preventing Chronic Disease* **1** 1-9.

Butterfield, R. M., Park, E. R., Puleo, E., Mertens, A., Gritz, E. R., Li, F. P., and Emmons, K. (2004). Multiple risk behaviors among smokers in the childhood cancer survivors cohort. *Psychooncology* **13**, 619-629.

Campbell, M. K., James, A., Hudson, M. A., Carr, C., Jackson, E., Oakes, V., Demissie, S., Farrell, D. and Tessaro, I. (2004). Improving multiple behaviors for colorectal cancer prevention among african american church members, *Health Psychol* **23**, 492-502.

Colditz, G. A., Cannuscio, C., and Frazier, A. (1997). Physical activity and reduced risk of colon cancer: Implications for prevention. *Cancer Causes and Control* **8**, 649-667.

Doll, R. and R. Peto (1981). The causes of cancer : quantitative estimates of avoidable risks of cancer in the United States today, *J. Natl. Cancer Inst.* **66**, 1191-1308.

Emmons, K. M., Marcus, B. H., Linnan, L. A., Rossi, J. S. and Abrams, D. B. (1999) Mechanisms in multiple risk factor interventions: Smoking, physical activity, and dietary fat intake among manufacturing workers. *Preventive Medicine* **23**, 481-489.

Emmons, K. M., Stoddard, A. M., Gutheil, C., Suarez, E. C., Lobb, R. and Fletcher, R. (2003). Cancer prevention for working class, multi-ethnic populations through health centers: The healthy directions study. *Cancer Causes and Control* **14**, 727-737.

Fairfield, K. M and Fletcher, R. H. (2002). Vitamins for chronic disease prevention in adults: scientific review. *Journal of the American Medical Association* **287**, 3116-3126.

Friedenreich, C. M., and Rohan, T. E. (1995). Physical activity and risk of breast cancer. *European Journal of Cancer Prevention* **4**, 145-151.

Giovannucci, E., Stampfer, M. J., Colditz, G., Hunter, D., Fuchs ,C., Rosner, B., Speizer, F. and Willett, W. (1998). Multivitamin use, folate, and colon cancer in women in the Nurse's Health Study. *Annals of Internal Medicine* **129**, 517-524.

Hosmer, D. W., and Lemeshow, S. (1998). *Encyclopeida of Biostatistics*, 2327-2333. John Wiley.

Hunt, M. K., Stoddard, A., Barbeau, E. M., Wallace, L. and Sorensen, G. (2003). Cancer prevention for working class, multiethnic populations through small businesses: The Healthy Directions Study. *Cancer Causes & Control* **14**, 749-760.

Institute of Medicine (2000). *Promoting Health: Intervention Strategies from Social and Behavioral Research.* National Academy Press.

Michaud, D. S., Augustsson, K., Rimm, E. B., Stampfer, M. J., Willett, W. C. and Giovannucci, E. (2001). A prospective study on intake of animal products and risk of prostate cancer. *Cancer Causes & Control* **12**, 557-567.

National Cancer Policy Board and Institute of Medicine (2003). Fulfilling the potential of cancer prevention and early detection, Technical Report, Washington D. C.

Prochaska, J. J. and 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**, 314-318.

Sandhu, M. S. and White, I. R. and McPherson, K. (2001). Systematic review of the prospective cohort studies on meat consumption and colorectal cancer risk: A meta-analytical approach. *Cancer Epidemiology, Biomarkers & Prevention* **10**, 439-446.

Vainio, H. and Bianchini, F. (2002). *IARC Handbooks of Cancer Prevention. Volume 6: Weight Control and Physical Activity.* IARC Press.

Received January 12, 2005; accepted June 15, 2005.

Melody S. Goodman
Graduate Program in Public Health
Department of Preventive Medicine, School of Medicine
SUNY, Stony Brook University
HSC Level 3, Rm 074
Stony Brook, NY 11794-8338
melody.goodman@stonybrook.edu

Yi Li
Department of Biostatistics and Computational Biology
Dana Farber Cancer Institute
44 Binney Street, M1B23, Boston, MA 02115
Yili@jimmy.harvard.edu

Gary G. Bennett
Center for Community Based Research
Dana Farber Cancer Institute
44 Binney Street, SM256, Boston, MA 02115
gary_bennett@dfci.harvard.edu

Anne M. Stoddard
Center for Statistical Analysis & Research
New England Research Institute
9 Galen Street Watertown, MA 02471
astoddard@neri.org

Karen M. Emmons
Center for Community Based Research
Dana Farber Cancer Institute
44 Binney Street, Boston, MA 02115
karen_m_emmons@dfci.harvard.edu