

Assessing the Effectiveness of Anti-smoking Media Campaigns by Recall and Rating Scores — A Pattern-Mixture GEE Model Approach

Ming Ji¹, Chengjie Xiong², Elizabeth A. Gilpin³ and Lois Biener⁴

¹San Diego State University, ²Washington University in St. Louis,

³University of California, San Diego, and ⁴University of Massachusetts

Abstract: Anti-smoking media campaign is an effective tobacco control strategy. How to identify what types of advertising messages are effective is important for maximizing the use of limited funding sources for such campaigns. In this paper, we propose a statistical modeling approach for systematically assessing the effectiveness of anti-smoking media campaigns based on ad recall rates and rating scores. This research is motivated by the need for evaluating youth responses to the Massachusetts Tobacco Control Program (MTCP) media campaign. Pattern-mixture GEE models are proposed to evaluate the impact of viewer and ads characteristics on ad recall rates and rating scores controlling for missing values, confounding and correlations in the data. A key difficulty for pattern-mixture modeling is that there were too many distinct missing data patterns which cause convergence problem for modeling fitting based on limited data. A heuristic argument based on collapsing missing data patterns is used to test the missing completely at random (MCAR) assumption in pattern-mixture GEE models. The proposed modeling approach and the recall-rating study design provide a complete system for identifying the most effective type of advertising messages.

Key words: Anti-smoking advertising, generalized estimating equations, missing completely at random, missing data, pattern-mixture models, rating score, recall rate, tobacco control.

1. Introduction

1.1 Anti-smoking media campaigns

Anti-smoking mass media campaign has been shown to significantly reduce the progression to regular smoking among both adults and adolescents, Flynn et. al (1992, 1994, 1995), Hu et. al (1995), Popham et. al (1995), Siegel et. al (1998, 2000) and Worden et. al (1996). However, development and evaluation of anti-smoking media campaign are difficult due to multiple reasons, such as insufficient

public or private funding, existence of multiple intervention programs and lack of evaluation methodology for design and analysis Worden *et. al* (1999). Many previous studies emphasize the behavioral changing effect of media campaigns on smoking initiation or cessation. Alternatively, another type of evaluation studies focus on the advertising effectiveness by assessing the relationship between ad characteristics and responses from the target audience. This type of studies addresses research questions such as “What type of ads are more likely to be remembered and are perceived to be more effectiveness?” These information are valuable for anti-smoking advertising agencies who are under the constraint of limited budget. There are relatively fewer studies of the second type in the literature. Goldman and Glantz (1998) studied 118 anti-tobacco advertisements and grouped them into thematic categories. They found that the messages on second-hand smoking and industry manipulation are effective in reaching all audiences compared to youth access to tobacco, short-term effects (such as yellow-teeth), long term health effects and romantic rejection. Pechmann and Goldberg (1998) used a message-based topology of large number of anti-smoking ads for youth to classify them into three types: fear appeals, peer norms and tobacco marketing manipulation. Comparison of the reaction from one set of a particular type of ads and a set of placebo ads (non-smoking related) shows that three subtypes (“smoking endangers the family unit”, “young smokers have taken the wrong path” and “most kids don’t smoke”) resulted in lower intentions to smoke. Biener, Keeler and Nyman (2000), Biener (2000) and Biener, Ming, Gilpin, and Alber (2004) presented data from in the Massachusetts Tobacco Control Program (MTCP) on the adult and youth response to eight different TV ads which were preclassified into three categories: sad, normative and funny. Their analysis concluded that the sad ads which evoke strong negative emotions had the highest recall rates and rating scores. This paper was motivated by the MTCP data analysis for developing a general statistical modeling approach for assessing anti-smoking media campaigns using recall and rating scores.

1.2 The MTCP program

The evaluation of the Massachusetts Tobacco Control Program (MTCP) began in 1993 with a monthly random-digit-dialed telephone survey containing questions on exposure to the anti-smoking mass media campaign and the recall of its tag line (“It’s time we made smoking history”). In 1997, the MTCP conducted a longitudinal follow-up survey of the teens interviewed at the 1993 baseline survey to evaluate reactions to the mass media campaign. Altogether there were 618 teens reinterviewed again in 1997. Below we describe the study design and measures used. The statistical methods are deferred to the METHODS section.

Respondents to the survey were asked if they could recall eight anti-smoking

Table 1: The eight ads in the Massachusetts Tobacco Control Program

Ad name	Brief description	Classification
1 Camel Ad	Shows a camel walking up to a microphone. He talks about how cartoon images of him have been used to sell cigarettes. He says it is outrageous.	Funny
2 Model Ad	Shows female soccer players talking about all the things they have been able to do without smoking cigarettes.	Normative
3 Janet Sackman Ad	Features a middle-aged women (Janet Sackman), who used to be a model in cigarette ads. She lost her vocal cords to cancer. She says she helped convince a lot of young people to smoke and now she wants to convince you not to.	Sad
4 Happy Birthday Ad	Shows a man who has a hole in his neck and has probably lost his vocal chords. He puts a microphone against his throat and sings Happy Birthday to the tobacco industry in a strange voice.	Sad
5 Pam Ad	Shows a young women, who started smoking so she could look older. When she was only 24 years old, she got very sick from smoking and had to have one of her lungs removed. Because of this, she has to take medicine for the rest of her life which makes her look a lot older than she really is.	Sad
6 Cowboy Ad	Shows a man who is the brother of the actor who played the Marlboro Man in cigarette ads. He talks about how tobacco companies use ads to make smokers look strong and independent. Then he talks about how his brohter got very sick and died from smoking.	Sad
7 Lung Ad	Shows a teenage boy having dinner at his girlfriend's home for the first time. The boy is a smoker. He is nervous and begins coughing. At the end of the commercial, he coughs up one of his lungs onto the dinner table.	Funny
8 Monica Ad	Shows a picture of a young woman getting sprayed with hair spray and splashed with mouthwash as she tries to get rid of a cigarette smell.	Normative

television ads, and for ads recalled, to rate each ad on a scale from 0 (the ad was not effective at all) to 10 (the ad was very effective). This design was based on the consideration of respondents' affective response to advertising which has been studied in marketing and consumer behavior research. The aim of the MTCP study was to assess two important dimensions of the impact of anti-smoking media campaigns: penetration and audience receptivity. These two dimensions reflect how successful the ads were in reaching the target audience and how effective the ads were in people's perceptions. Quantitatively, these dimensions were measured by the recall rate and the rating score for each ad queried in the survey. These two measures may be affected by the characteristics of both the ads and the viewers. Viewer characteristics included demographic variables, TV viewing habits, household education level, ownership of tobacco promotional items and other factors. There were eight ads in the survey: Camel,

Model, Janet Sackman, Happy Birthday, Pam, Cowboy, Lung and Monica. The characteristics of the eight ads were first screened and rated by an independent panel of judges (104 young people) and finally categorized into 3 groups: (1) sad/frightening ads (2) normative ads, and (3) funny ads. Table 1 contains the description and the classification of the 8 ads surveyed in the MTCP. For more background information about this study, please refer to Biener (2000), which also contains an analysis of the youth’s responses to the eight TV ads. The main finding in Biener (2000) was that sad ads, which provoke strong negative emotions, were remembered the most and rated the highest in effectiveness.

1.3 Organization of the paper

In this paper, we present a statistical methodology for assessing the effectiveness of anti-smoking media campaigns based on the MTCP study. The structure of this paper is as follows. In the METHOD Section, we discuss three intrinsic statistical issues related to the MTCP data and the rationale for using GEE pattern-mixture models for the analysis. In the DATA ANALYSIS Section, we discuss the implementation of our method and present analysis results and interpretation. In the Conclusion and Discussion Section, we discuss both the strengths and the limitations of our analysis and point out some possible directions for future research efforts.

2. Method

2.1 Statistical issues in data analysis

In this section, we discuss several challenging issues intrinsic to the analysis of the MTCP data. First, there were significant amount of missing values in the rating scores since not many people had seen or remembered seeing all eight ads. Second, the rating scores are cognitive measurements from the same person and therefore likely to be correlated. Third, both viewer and ad characteristics may confound ad recall and rating. Below, we discuss these three issues in details and outline ways to handle them.

Missing Data Let y_{ij} be the rating score by the i -th subject for the j -th ad, then the raw rating scores from all the 618 teens on 8 ads in the study form a 618×8 data matrix $\mathbf{Y} = (y_{ij})$, where $i = 1, \dots, 618, j = 1, \dots, 8$. From the data matrix \mathbf{Y} , we determined that there were 3051 scores observed out of $618 \times 8 = 4944$ possible scores. The percentage of missing values in the rating scores is high at about 38%. This is because not many people would see or remember all eight ads.

Let $m_{ij} = 0$ if y_{ij} is missing and $m_{ij} = 1$ otherwise. Then, corresponding to

the data matrix \mathbf{Y} , we have a 618×8 matrix $\mathbf{M} = (m_{ij})$ whose entities are either 0 or 1. is usually called the missing data matrix in the statistical literature.

The rows of \mathbf{Y} and \mathbf{M} are rating scores and the recall indicators, respectively. It is easy to compute the mean rating score or the mean recall rate for each ad by averaging the columns of \mathbf{Y} and \mathbf{M} . However, it is well known that this approach may lead to biased estimates when the missing data are not Missing Completely at Random, i.e., MCAR (Little and Rubin (1987)).

Confounding Factors The relationship between the rating score and potentially confounding factors is fairly complex. For example, the intensity of advertising or frequency of ad airing for each ad may affect recall; the higher the intensity, the higher the recall rate might be. Smoking status may affect ratings because established smokers may have negative attitude toward anti-smoking ads and rate them lower. In the MTCP study, there were a variety of potentially confounding factors measured on demographics, TV viewing habits and ad characteristics. In order to obtain accurate estimates of the rating scores and recall rates, adjustment for these potential confounding factors is necessary.

Correlated Data The third issue is the correlation among the outcome measures. The rating scores y_{ij} as well as the recall indicators m_{ij} , for fixed i , are repeated measures from the same person and therefore correlated. For example, a person's perception of sad ads may be similar but could be very different from his perception of normative ads. As a result, this person might rate all sad ads higher than all normative ads. However, although this correlation exists, it is not our primary concern. Our goal is to obtain an accurate estimate of the relationship between ad characteristics and viewer characteristics and the recall and rating scores. By taking into account the correlations among the outcomes, we can develop more efficient estimates for recalls and rating scores.

1.2 Statistical modeling approach

Pattern-Mixture Models with Covariates Let \mathbf{x} be a vector of covariates that summarize the viewer and ad characteristics. Let \mathbf{y} and \mathbf{m} denote the generic random vector that generates the rows of \mathbf{Y} and \mathbf{M} , respectively. Then the goal of controlling confounding and missing data can be achieved by estimating the regression between \mathbf{x} and (\mathbf{y}, \mathbf{m})

The conditional distribution of \mathbf{y} and \mathbf{m} given \mathbf{x} can be decomposed as follows

$$f(\mathbf{y}, \mathbf{m}|\mathbf{x}) = f(\mathbf{m}|\mathbf{x})f(\mathbf{y}|\mathbf{m}, \mathbf{x}) \quad (2.1)$$

This decomposition is named *the pattern-mixture model with covariates* by Little and Wang (1996). It is proposed as an alternative approach of *the selection-model*

with covariates for the missing data problem. The selection model approach has the following decomposition

$$f(\mathbf{y}, \mathbf{m}|\mathbf{x}) = f(\mathbf{m}|\mathbf{y}, \mathbf{x})f(\mathbf{y}|\mathbf{x}) \quad (2.2)$$

Note that in (2.1), $f(\mathbf{y}|\mathbf{m}, \mathbf{x})$ is a mixture over distinct missing data patterns. The idea of the pattern-mixture modeling approach to missing data is to stratify the study sample based on missing data patterns, without directly modeling the missing data mechanism as in the selection model approach. More discussions on the pattern-mixture model approach to missing data problems can be found in Little and Wang (1996), Little (1993) and Little (1994).

The pattern-mixture model with covariates decomposition (2.1) is a complete solution for our purpose of modeling the recall rates and the rating scores as functions of viewer and ad characteristics. Note that $f(\mathbf{m}|\mathbf{x})$ describes how the viewer and ads characteristics affect the recalls. It is exactly the model for the recall of the anti-smoking advertising. It can answer questions such as: "What type of ads are more likely to be remembered?" On the other hand, modeling $f(\mathbf{y}|\mathbf{x})$ addresses the perceived effectiveness of the ads. It can answer questions such as: "what type of ads are rated higher, (i.e., perceived to be more effective)?" Although we have $f(\mathbf{y}|\mathbf{m}, \mathbf{x})$ in (2.1), we will show that our data supports the hypothesis $H_0 : f(\mathbf{y}|\mathbf{m}, \mathbf{x}) = f(\mathbf{y}|\mathbf{x})$ using a test developed by Park and Lee (1997) for testing MCAR in repeated measures data. The meaning of H_0 is that the regression between \mathbf{y} and \mathbf{x} is homogeneous across different missing data patterns. It is equivalent to $\mathbf{y} \perp \mathbf{m}|\mathbf{x}$, where \perp means "independent of", which is MCAR in the traditional Rubin and Little (1987) sense.

Generalized Estimating Equations Since the recalls and rating scores on different ads from the same person are correlated, statistical models for repeated measures that take into account of this correlation will provide more efficient estimates than fitting one model for each ad. There is an extensive statistical literature for modeling repeated measures data as described in Diggle, Zeger and Liang (1994). In particular, generalized estimating equations (GEE) are a class of repeated measures models. Several fundamental papers on GEE are Liang and Zeger (1987), Zieger and Liang (1986), Zieger and Liang (1988) and Prentice (1988). An excellent review of GEE theory can be found in Ziegler, Kastner, Gromping, and Blettner (1996). GEE is implemented in many statistical packages such as PROC GENMOD in SAS. A review on GEE software is Horton and Lipsitz (1999).

To be more specific, denote the repeated measures as y_{ij} , which is the j -th measurements for i -th subject. Let $\mathbf{y}_i = (y_{i1}, \dots, y_{it_i})^T$ be the $t_i \times 1$ ($t_i \leq t$) vector of responses and $\mathbf{x}_i = (x_{i1}, \dots, x_{it_i})^T$ the $t_i \times p$ matrix of covariate values, where

x_{ij} is a $1 \times p$ vector of covariates for the i -th subject. Let us assume the marginal distribution of y_i is in the exponential family

$$f(y_{ij} = \exp \left[\frac{y_{ij}\theta_{ij} - b(\theta_{ij})}{a(\phi)} + c(y_{ij}, \phi) \right]) \quad (2.3)$$

where a, b and c are specific functions with canonical parameters θ_{ij} and scale parameter ϕ Zeger and Liang (1988). The mean and variance are assumed to be

$$E(y_{ij}) = \mu_{ij} = b'(\theta_{ij}) \quad (2.4)$$

$$Var(y_{ij}) = b''(\theta_{ij})a(\phi) \quad (2.5)$$

respectively. The relationship between the response and covariates is described by the following regression model

$$\boldsymbol{\eta}_i = \mathbf{x}_i\boldsymbol{\beta} \quad (2.6)$$

where $\boldsymbol{\eta}_i = (\eta_{i1}, \dots, \eta_{it_i})^T$ with $\boldsymbol{\eta}_{ij} = g(\boldsymbol{\mu}_{ij}) = \mathbf{x}_{ij}\boldsymbol{\beta}$. Here $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ is the vector of the unknown regression parameters to be estimated and g is known as the link function. Liang and Zeger showed that the GEE estimate for $\boldsymbol{\beta}$ is consistent and asymptotically normal if the mean structure (2.4) is correctly specified, even if the covariance structure $Var(\mathbf{y}_i)$ is misspecified Liang and Zeger (1987) and Zeger and Liang (1988). This property makes GEE very suitable for our analysis, because our goal is to estimate the mean regression between the rating scores and the confounding variables while accurate information about the correlation structure is not a priority.

Note that for the GEE models, any arrangement of the second index j represents a fixed order of the repeated measures. In our data, the ratings scores are arranged according to their natural order in the survey. Ideally, the order should be randomized, because the recall and rating could be affected by the order of the ads. The effect of the ordering in the current MTCP data cannot be evaluated because no randomization of ordering was done. This is a limitation of this study, and further discussion on this issue is in the conclusion and discussion Section.

Testing MCAR As we point out in previous sections, a prerequisite for applying the GEE pattern-mixture model for the rating scores is to show that the missing values are MCAR as formalized in $H_0 : f(\mathbf{y}|\mathbf{m}, \mathbf{x}) = f(\mathbf{y}|\mathbf{x})$. In this subsection, we describe a method from Park and Lee (1997) to test the hypothesis H_0 . Their idea is to derive an extended GEE which includes an additional term corresponding to missing patterns and then test if the regression parameters for this term are significantly different from 0. The following description of their approach is adopted from Park and Lee (1997). In the presence of missing data, let us assume that there are K distinct missing data patterns. Let S_k denote the set

of observations with missing data pattern k ($k = 1, 2, \dots, K$). Define indicator variables I_{ik} so that $I_{ik} = 1$ if the i -th observation is in S_k and $I_{ik} = 0$ otherwise. Using these definitions, Park and Lee (1997) showed that the MCAR hypothesis $H_0 : f(\mathbf{y}|\mathbf{m}, \mathbf{x}) = f(\mathbf{y}|\mathbf{x})$ can be tested by fitting the following extended GEE model:

$$\boldsymbol{\eta}_i = (\mathbf{1} \otimes \mathbf{I}_i) \boldsymbol{\alpha} + \mathbf{x}_i \boldsymbol{\beta} \quad (2.7)$$

where $\mathbf{1}_i$ is a $t_i \times 1$ vector of 1's, $\mathbf{I}_i = (I_{i1}, \dots, I_{iK})$ and $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K)^T$. The hypothesis H_0 is equivalent to

$$H_0^* : \alpha_1 = \dots = \alpha_K = 0 \quad (2.8)$$

If H_0^* does not hold, then the η_{ij} 's, which are the overall mean values of the y_{ij} 's differ across the S_k 's. This means that the missing data affects the regression between the response \mathbf{y} and the covariates \mathbf{x} .

Let $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ denote the GEE estimates for (2.7) and $\mathbf{V}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}$ the asymptotic variance-covariance matrix. Let $\mathbf{V}_{\boldsymbol{\alpha}}$ be the submatrix of $\mathbf{V}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}$ corresponding to $\boldsymbol{\alpha}$. A generalized Wald test statistic for H_0^* is $T_w = n^{-1} \hat{\boldsymbol{\alpha}}^T \hat{\mathbf{V}}_{\hat{\boldsymbol{\alpha}}} \hat{\boldsymbol{\alpha}}$, where $\hat{\mathbf{V}}_{\hat{\boldsymbol{\alpha}}}$ is the "robust" estimator of the variance-covariance matrix associated with $\boldsymbol{\alpha}$. T_w has an approximate chi-square distribution with K degrees of freedom under H_0^* . Actually, $T_w = n^{-1} \hat{\boldsymbol{\alpha}}^T \hat{\mathbf{V}}_{\hat{\boldsymbol{\alpha}}} \hat{\boldsymbol{\alpha}}$ is exactly the generalized Wald test statistic in Rotnitzky and Jewell (1990). The generalized Wald test statistic is available in SAS PROC GENMOD by specifying the options TYPE3 and WALD in the MODEL statement. Another related test statistic in GEE is the "score" test statistic which is the default of the TYPE3 option in PROC GENMOD. Both can be used to test the significance of an effect in a GEE model. Therefore, the Park and Lee test can be implemented easily using SAS PROC GENMOD.

A potential problem with this approach occurs when the number of distinct missing patterns K is large. This is because pattern-mixture modeling essentially requires stratification of the sample according to distinct missing data patterns. When K increases, the number of strata increases and the number of unknown parameters for the GEE model (2.7) also increases rapidly. Unless the sample size is sufficiently large, there will be some strata with sparse data that cause convergence problems for the GEE model fitting. This is a major difficulty for applying pattern-mixture models to missing data problems. To our knowledge, currently there is no sound statistical methodology to handle this problem. In this paper, we propose a heuristic approach based on collapsing the distinct missing data patterns to apply the Park and Lee's method. Our idea is to collapse the missing data patterns into a smaller number of patterns and run the Park and Lee test, and then to repeat this procedure many times according to different ways of collapsing the missing data patterns. If the MCAR is consistently rejected or

Table 2: Explanatory variables

Explanatory variables	Definition	Values
Demographics		
GENDER	gender	0-female 1-male
AGE	age	15-21
RACE	race	0 - minority 1-white
TV Viewing Habit		
B10	weekday hours watching TV	0-none 1-1hour or less 2-2 to 24 hours
B15	Saturday hours watching TV	0-none 1-1hour or less 2-2 to 24 hours
Smoking Status		
SMK	smoking status	1-nonsmokers 2-susceptable nonsmokers and experimenter 3-established smoker
Household		
HHED	education of head of household	0 - more than high school 1 - high school
HSMK	at least one adult smoker in household	0 - no 1 - yes
FRNDSMK	one or more smoker in household smoker in household	0 - no 1 - yes
Own Tobacco Item		
OWN	own tobacco pro- motional items	0 - no 1-yes
Ad Charateristics		
ADTYPE	type of ad	0 - normative 1 - funny 2 - sad
REM	number of months since an ad last aired	continuous
GRP	intensity of advertising	continuous

accepted across different scenarios of collapsing, then it is plausible to reject or accept MCAR. More details related to this issue are in the DATA ANALYSIS section.

To summarize the discussion of this section, we will fit two GEE models for $f(m|x)$ and $f(y|x)$ for assessing the recall and rating of the ads in our study after the Park and Lee test confirms that hypothesis $H_0 : f(\mathbf{y}|\mathbf{m}, \mathbf{x}) = f(\mathbf{y}|\mathbf{x})$ is plausible.

3. Data Analysis

3.1 Implementation

We implemented the GEE analysis proposed in the previous section using PROC GENMOD in SAS 8.1. PROC GENMOD contains several useful functionalities that suit our purpose. For example, besides the utility of the generalized TYPE 3 Wald statistic, the CONTRAST and LSMEANS statements in PROC GENMOD allow us to compare different levels of categorical variables, especially for interaction terms as shown in our subsequent analysis.

Dependent and Independent Variables The dependent variables are the recall indicators and rating scores, respectively. The independent variables used are listed in Table 2. They describe the characteristics of the viewers and the ads.

Collapsing Missing Data Patterns As pointed out previously, a large number of missing data patterns can cause convergence problem for the pattern-mixture modeling. Since there are eight ads, there are $2^8 - 1 = 255$ possible missing data patterns. For example, let "M" represent that the observation is missing and "O" represent that it is observed, then OOOOOOOO represents all eight ads were recalled and rated; OOOOOOOM represents the first seven ads were rated but the rating for the last ad was missing. The 255 possible missing data patterns will create many strata with sparse data because there were only 618 study participants. As mentioned before, we will use different schemes to collapse the missing data patterns.

For example, we can use the number of ads recalled to present the missing data patterns, see the first case of Table 3. There are nine of these patterns but the pattern of missing all of the rating scores can be excluded since there is no response and the individual contribution to the estimating equation cannot be constructed. Thus, the missing data pattern variable, is defined as the subset of those who have k observed rating scores, where $k = 1, \dots, 8$.

Table 3: p -values for different missing data patterns

K	S_k 's	p
8	$S_k = \{\text{no. of missing rating scores}\}$	0.63
9	$S_1 = \{OOOOOOOO\}$ $S_2 = \{OOOOOOOM\}$ $S_8 = \{OMMMMMMM\}$ $S_9 = \{\text{others}\}$	0.07
9	$S_1 = \{MOOOOOOO\}$ $S_2 = \{OMOOOOOO\}$ $S_8 = \{OOOOOOOM\}$ $S_9 = \{\text{others}\}$	0.24
3	$S_1 = \{\text{missing one of the first four rating scores}\}$ $S_2 = \{\text{missing one of the last four rating scores}\}$ $S_3 = \{\text{others}\}$	0.56

Collapsing missing data patterns has another problem. For the S_k 's specified above, for example, S_6 may include the missing data pattern MMOOOOOO or OOOOOO MM among others. So after the collapsing, there are still different missing data patterns within each S_k . From GEE theory, even with missing values in the response variable, GEE may still lead to consistent estimates for the regression parameters, but only when the missing values are MCAR, Liang and Zeger (1987). Therefore the formulation of the S_k 's based on number of missing values requires an assumption that within each S_k , the missing data are MCAR. Under this assumption, the GEE model will lead to consistent estimate of (α_k, β) in each S_k for testing H_0^* . This is a strong assumption that we have to make for our current analysis.

A remedy for this strong assumption on collapsing missing data patterns is to exploit the flexibility in formulating the S_k 's in Park and Lee's approach. For example, we can define S_k as the group of all the subjects who are "drop-outs" from the k -th position, where $k = 0, 1, \dots, 8$. Then we will have ten strata, namely, S_0, S_1, \dots, S_8 and S_9 , which consists of all the observations that do not have a "monotone" drop-out pattern. If we experiment with different ways of collapsing missing data patterns in terms of different ways of formulating S_k 's for smaller $K (< 256)$ and H_0^* always holds, then it is more defensible that H_0^* holds for $K = 256$. We have listed several different ways of collapsing the missing data pattern in Table 3. The p -values from the score test of GEE for H_0^* will be used

Table 4: GEE model fitting for rating scores

Explanatory Variables	Final GEE Model		
	Estimate	Std. Error	p -value ^a
Intercept	-0.06	0.10	< 0.0001
Gender			< 0.0001
female	0.21	0.09	
male	0.00	0.00	
ADTYPE			< 0.0001
normative	-0.12	0.10	
funny	-1.59	0.11	
sad	0.00	0.00	
REM	-0.01	0.004	0.003
GRP	0.08	0.006	< 0.0001
AGRGRP			0.0009
12<AGE<13	0.26	0.08	
14<AGE<14	0.00	0.00	
GENDER*ADTYPE			0.0099
(female, normative ads)	-0.01	0.14	
(female, funny ads)	0.41	0.14	
(female, sad ads)	0.00	0.00	
(male, normative ads)	0.00	0.00	
(male, funny ads)	0.00	0.00	
(male, sad ads)	0.00	0.00	

^aThe p -value is from type 3 Wald test statistic except the p -value for the intercept which is from a two-sided t -test.

to test the significance of the missing data pattern. We will conclude that it is reasonable to assume MCAR for the data, if the missing data patterns are not significant no matter how we collapse. This argument is a heuristic solution for treating the difficult problem of having too many distinct missing data patterns in pattern-mixture modeling. A more rigorous methodology for treating this missing data pattern problem still needs to be developed.

Correlation Structure To fit a GEE model, we also must specify a correlation structure for the repeated measurements, in our case, the correlation among the rating scores. Although we may hypothesize that similar ads may be rated similarly, we do not have a precise understanding of this correlation. This discourages the use of a particular correlation structure such as $AR(1)$. To be as general as possible, we used the unstructured correlation matrix in our GEE analysis. Fortunately, even with this general correlation structure, the GEE models

Table 5: GEE model fitting for rating scores

Explanatory Variables	Final GEE Model		
	Estimate	Std. Error	<i>p</i> -value ^a
Intercept	2.74	0.18	<0.0001
AGEGRP	–	–	–
GENDER			0.0001
female	0.62	0.13	
male	0.00	0.00	
RACE			0.0003
minority	0.54	0.14	
white	0.00	0.00	
OWN			0.015
no	0.40	0.16	
yes	0.00	0.00	
ADTYPE			<0.0001
normative	–1.28	0.13	
funny	–1.38	0.12	
sad	0.00	0.00	
GRP	–0.03	0.007	0.0003
GENDER*ADTYPE			0.034
(female, normative ads)	–0.45	0.19	
(female, funny ads)	0.14	0.17	
(female, sad ads)	0.00	0.00	
(male, normative ads)	0.00	0.00	
(male, funny ads)	0.00	0.00	
(male, sad ads)	0.00	0.00	

^aAll the *p*-values are from type 3 Wald test statistic except the *p*-value for the intercept which is from two-sided *t*-test.

converged quickly. This is also a benefit of reducing the number of missing data patterns.

3.2 Results and interpretation

Model Fitting For both recall rates and rating scores, we fit several GEE models. The first GEE model was fitted using all the explanatory variables in Table 2 and two interaction terms, GENDER*ADTYPE and AGEGRP*ADTYPE. In order to form the interaction term, we dichotomized AGE into AGEGRP by defining AGEGRP to be 0 if AGE is between 12 and 13 or 1 if AGE is between 14 and 15. For rating scores, an additional term representing the missing

data pattern MISSPATN is included. Next, another GEE used the significant or marginally significant variables from the first GEE based on the p -values from the generalized Wald statistics of TYPE 3 analysis ($p < 0.15$). The final GEE model included only significant predictors. For the recall rates, 3 GEE models were fitted, and 2 were fitted for the rating scores.

Significant Variables The resulting estimates from the final GEE models are listed in Tables 4 and 5 for the recall rates and the rating scores, respectively.

In Table 4, the p -values for the parameter estimates show that only the INTERCEPT, AGE, GENDER, GRP (advertising intensity) and ADTYPE (type of ads) remained significant in the final GEE model. Sad ads were more likely to be recalled. Females remembered more ads. The longer time since ads aired last time, the less likely they would be remembered. Younger adolescents recalled more ads. The hours of TV watching (B10, B15), smoking status (SMK), social influences (HHED, HHSMK, FRNDSMK), and ownership of tobacco promotional items (OWN) did not significantly affect recall rates.

For Table 5, we started to use the missing data pattern in which equal numbers of ads recalled. The p -value associated with MISSPATN from the TYPE 3 generalized Wald statistic was 0.63. We then tried different sets of strata, S_k as specified in Table 3. We observed that H_0^* held across different scenarios. All the p -values were larger than 0.05 for the different ways of collapsing the missing data patterns shown in Table 3. This suggests that our data strongly support H_0^* . Therefore the hypothesis $H_0 : f(\mathbf{y}|\mathbf{x}, \mathbf{m}) = f(\mathbf{y}|\mathbf{x})$ is a plausible assumption in the application of the pattern-mixture model for this data set. The final GEE model for the rating scores in Table 5 shows that the only significant effects on the rating scores were: the INTERCEPT, RACE, GENDER, OWN, GRP and ADTYPE. Sad ads were the perceived to be most effective. Female adolescents rated the ads higher than their male counterparts. Minority teens rated ads higher than white kids. Owning tobacco promotional items was negatively associated with the rating scores.

4. Conclusion and Discussion

We have presented a statistical modeling approach to evaluate the effectiveness of anti-smoking media campaigns according to market penetration and perceived effectiveness of the ads as measured by ad recall and rating. The fitted models provide more efficient estimates of recall rates and rating scores, controlling for missing values, confounding variables and correlations among these outcome measures. The modeling approach offers a flexible way to identify significant viewer characteristics and ad type and execution elements on the effectiveness of ads.

Our approach also has several limitations. The first is a design issue: the order of the ads was not randomized and whether or how this will influence our results is unclear. The order of the ads asked may have an effect on recall or rating scores. For example, repeatedly asking about several sad ads together instead of mixing them with other types of ads could strengthen a nonsmoker's memory of the negative emotions associated with sad ads and lead to higher recall or rating. A smoker, however, may object to being repeatedly reminded of sad anti-smoking messages and give lower ratings. The ads surveyed in the MTCP were arranged in a fixed order. Therefore, the potential bias caused by this fixed order of ads cannot be evaluated. Randomizing ad order should be considered in future designs. The second limitation is that we used the ADTYPE variable for all the ads in the same category. This is a simplification, which implies homogeneity of recall and rating scores across different ads within the same categories. However, we can solve this problem by dropping the ADTYPE variable and using the ad index, namely, j ($= 1, 2, 3, \dots, 8$), in the GEE models. This will lead to ad-specific recall rates and rating scores for each individual ad. The third limitation is the heuristic approach to the large number of missing data patterns. More rigorous methods are needed for this difficult issue in applying pattern-mixture modeling.

Despite the limitations, our modeling approach, combined with design based on recall and rating would provide a systematic way for assessing the effectiveness of advertising messages in general. For example, the data collection of recall and rating can be conducted in more controlled laboratory settings. We can show different types of advertising messages to study subjects and then ask them to recall and rate each message after a short time (such as next day). The rating scores from the same person on multiple ads can be analyzed using the GEE modeling approach and the amount of missing data will be much reduced due to the short follow-up time. Therefore, we can evaluate the effectiveness of different advertising messages in a laboratory environment before putting them in mass media.

Acknowledgement

Dr. Ming Ji was supported by a two-year Research Fellowship from Cancer Research Foundation of America while conducting this research. Ms Gilpin's contribution to this project was supported by Grant #9RT-W36 from the University of California Tobacco Related Disease Research Program.

References

- Biener, L., Keeler, G. M. and Nyman, A. L. (2000). Adults' response to Massachusetts's antitobacco media campaign: Impact of viewer and ad characteristics. —*it Tobacco Control* **9**, 401-407.

- Biener, L., Ji, M., Gilpin, E. and Alber, A. (2004). The impact of emotional tone, message and broadcast parameters in youth anti-smoking advertisement. *Journal of Health Communication* **9**, 259-274.
- Biener, L. (2000) Commentary: Adult and youth response to the Massachusetts anti-tobacco television campaign. *Journal of Public Health Management and Practice* **6**, 40-44.
- Flynn, B. S. *et al.* (1992). Prevention of cigarette smoking through mass media intervention and school programs. *American Journal of Public Health* **82**, 827-834.
- Flynn, B. S. *et al.* (1994). Mass media and school interventions for cigarette smoking prevention: Effects 2 years after completion. *American Journal of Public Health* **84**, 1148-1150.
- Flynn, B. S. *et al.* (1995). Cigarette smoking prevention effects of mass media and school interventions targeted to gender and age groups. *Journal of Health Education* **26**, 45-51.
- Goldman, K. L. and Glantz, S. A. (1998). Evaluation of antismoking advertising campaigns. *Journal American Medical Association* **279**, 772-777.
- Horton, N. J. and Lipsitz, S. R. (1999). Review of software to fit generalized estimating equation regression models. *The American Statistician* **53**, 160-169.
- Hu, T. W., Sung, H. Y. and Keeler, T. E. (1995). Reducing cigarette, 0 consumption in California — tobacco taxes vs anti-smoking media campaign. *American Journal of Public Health* **85**, 1218-1222.
- Liang, K. Y., and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models *Biometrika* **73**, 13-22.
- Little, R. J. A., and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. Wiley.
- Little, R. J. A., and Wang, Y. X. (1996). Pattern-mixture models for multivariate incomplete data with covariates. *Biometrics* **52** , 98-111.
- Little, R. J. A. (1993). Pattern-Mixture models for multivariate incomplete data. *Journal of American Statistical Association* **88**, 125-134.
- Little, R. J. A. (1994). A class of pattern-mixture models for normal incomplete data. *Biometrika* **81**, 471-483.
- Park, T., and Lee, S. Y. (1997). A test of missing completely at random for longitudinal data with missing observations. *Statistics in Medicine* **16**, 1859-1871.
- Peter, J. D., Liang, K. Y., and Zeger, S. L. (1994). *Analysis of Longitudinal Data*. Oxford Science Publication.
- Pechman, C. and Goldberg, M. (1998). *Evaluation of Ad Strategies for Preventing Youth Tobacco Use*. 1998 report submitted to the California Tobacco Related Disease Research Program.
- Popham, W. T. *et al.* (1993). Do Anti-smoking media campaign help smokers quit? *Public Health Reports* **108**, 510-513.

- Prentice, R. L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics* **44**, 1033-48.
- Rotnitzky, A., and Jewell, N. P. (1990). Hypothesis testing of regression parameters semiparametric generalized linear models for clustered correlated data. *Biometrika* **77**, 485-497.
- Siegel M. (1998). Mass media anti-smoking campaign: a powerful tool for health promotion. *Annals of Internal Medicine* **129**, 128-132.
- Siegel M. and Biener L. (2000). The impact of antismoking media campaign on progression to established smoking: results of a longitudinal youth study. *American Journal of Public Health* **90**, 380-386.
- Worden, J. K. *et al.* (1996). Using mass media to prevent cigarette smoking among adolescent girls. *Health Education Quarterly* **23**, 453-468.
- Worden, J. K. (1999). Research in using mass media to prevent smoking. *Nicotine and Tobacco Research* **1**, S117-S121.
- Zeger, S. L., and Liang, K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* **42**, 121-130.
- Zeger, S. L., and Liang, K. Y. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics* **44**, 1049-1060.
- Ziegler A., Kastner C., Gromping U., and Blettner, M. (1996). The generalized estimating equations in the past ten years: an overview and a biomedical application. *Biometrical Journal* **40**, 115-139.

Received October 16, 2005; accepted December 9, 2005.

Ming Ji
Division of Epidemiology and Biostatistics
Graduate School of Public Health
San Diego State University
5500 Campanile Drive
San Diego, CA 92128-4162, USA
mji@mail.sdsu.edu

Chengjie Xiong
Division of Biostatistics
Washington University in St. Louis
St. Louis, MO63110, USA
chengjie@wubios.wustl.edu

Elizabeth A. Gilpin
Department of Family and Preventive Medicine
University of California, San Diego
La Jolla, CA 92093-0645, USA

bgilpin@ucsd.edu

Lois Biener
Center for Survey Research
University of Massachusetts
Boston, MA 02125, USA
lois.biener@umb.edu