



Published in final edited form as:

Inf Sci (Ny). 2010 February 1; 180(3): 432–440. doi:10.1016/j.ins.2009.09.018.

Estimation of Transitional Probabilities of Discrete Event Systems from Cross-Sectional Survey and its Application in Tobacco Control

Feng Lin¹ and Xinguang Chen²

¹Department of Electrical and Computer Engineering, Wayne State University, Detroit, MI 48202, USA and School of Electronics and Information Engineering, Tongji University, Shanghai, China. flin@ece.eng.wayne.edu, Tel: 313 5773428, Fax: 313 5785844

²Pediatric Prevention Research Center, Wayne State University, Detroit, MI 48202, USA. jimchen@med.wayne.edu

Abstract

In order to find better strategies for tobacco control, it is often critical to know the transitional probabilities among various stages of tobacco use. Traditionally, such probabilities are estimated by analyzing data from longitudinal surveys that are often time-consuming and expensive to conduct. Since cross-sectional surveys are much easier to conduct, it will be much more practical and useful to estimate transitional probabilities from cross-sectional survey data if possible. However, no previous research has attempted to do this. In this paper, we propose a method to estimate transitional probabilities from cross-sectional survey data. The method is novel and is based on a discrete event system framework. In particular, we introduce state probabilities and transitional probabilities to conventional discrete event system models. We derive various equations that can be used to estimate the transitional probabilities. We test the method using cross-sectional data of the National Survey on Drug Use and Health. The estimated transitional probabilities can be used in predicting the future smoking behavior for decision-making, planning and evaluation of various tobacco control programs. The method also allows a sensitivity analysis that can be used to find the most effective way of tobacco control. Since there are much more cross-sectional survey data in existence than longitudinal ones, the impact of this new method is expected to be significant.

Keywords

Discrete event systems; tobacco control; smoking behavior; cross-sectional survey; transitional probability

I. INTRODUCTION

Reducing tobacco use remains a significant public health challenge in the new millennium despite decades of efforts in tobacco control. Exposure to tobacco is associated with 440,000 deaths each year in the United States and costs the nation \$50 – \$75 billion in medical expense alone [1–9]. The US government aims at reducing current smokers to 16% among adolescents

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

and 12% among adults by the year 2010 [2]. However data from the Behavioral Risk Factor Surveillance System indicate that the adult smoking rate has fluctuated at around 22% since 1990 [1–9]. Data from the Youth Risk Behavior Surveillance System indicate that adolescent smoking prevalence rate increased from 27.5% in 1991 to 36.4% in 1997 before it started to decline in 1999 [1–9]. Hence, there is a large gap between the current levels of tobacco use and the tobacco control objective, underscoring the need for immediate actions to advance current tobacco control strategy.

Effective tobacco control strategies require a comprehensive understanding of the dynamics of smoking behavior progression. The continuous development of smoking behavior can be effectively modeled as a discrete event system (DES), where discrete states describe the different stages of tobacco use and discrete events describe the transitions from one state (stage) to another. Much has been documented on various types of smokers (e.g., smoking experimenters, regular smokers, addicted smokers, quitters etc), reflecting the stages of smoking. However, data are needed on the transitional (or event) probabilities among various stages of tobacco use. Traditionally, such probabilities are estimated by performing longitudinal surveys; that is, by surveying individuals of the same group of randomly sampled subjects over time. Longitudinal surveys, often time-consuming and expensive, are less frequently performed. On the other hand, cross-sectional surveys, where random samples of a population are surveyed at the same time are much easy to perform compared to longitudinal surveys. Hence many more cross-sectional surveys have been conducted, resulting in abundant cross-sectional data. Therefore the challenge question is: can we estimate the transitional probabilities from cross-sectional survey data? Intuitively, one may think that the answer shall be negative. But, in fact, as we will show in this paper that there is information contained in cross-sectional surveys that allows us to derive the transitional probabilities. We will develop a new approach to estimate the transitional probabilities from cross-sectional survey data using a new framework of discrete event systems. Extracting transitional probabilities from cross-sectional survey data has never been attempted before. We are the first to suggest this possibility.

To estimate transitional probabilities from cross-sectional survey data, we will extend the conventional framework of discrete event systems by introducing state probabilities and transitional probabilities. To establish the DES method for modeling smoking behavior, we use the following five states:

- NS – nerve-smoker, a person who has never smoked by the time of the survey.
- EX – experimenter, a person who smokes but not on a regular basis after initiation.
- SS – self stopper, an ex-experimenter who stopped smoking for at least 12 months.
- RS – regular smoker, a smoker who smokes on a daily or regular basis (including habitual and addicted smokers).
- QU – quitter, an ex-regular smoker who stopped smoking for at least 12 months.

We will establish a set of equations that relate the probabilities of the states defined above with the transitional probabilities based on data available in cross-sectional surveys. Using probabilistic discrete event system (PDES) to study smoking behavior is new and our approach is the first of its kind.

The transitional probabilities describe the dynamics of smoking behavior progression in a population. They can also be used to predict future trend of tobacco use as well as the effectiveness of tobacco control programs at national and state levels. By associating various tobacco control programs with changes in transitional probabilities, we can determine the

impact of these programs in reducing tobacco use and obtain useful information for tobacco control planning in the future.

The results presented in this paper are original and important. (1) We propose a PDES model for smoking behavior, which has not been previously attempted. This model is very intuitive and easy to understand. (2) We establish a method to estimate transitional probabilities from cross-sectional survey data, which has been commonly considered impossible. (3) We use the proposed estimation method on the actual survey data and obtain the transitional probabilities from the data for the first time. (4) We use the estimated transitional probabilities to predict the future smoking behavior, which provides a new tool in tobacco control. (5) We develop sensitivity analysis using PDES model for smoking behavior, which has never been done before.

This paper is organized as follows. In Section II, we will discuss advantages and disadvantages of cross-sectional surveys and longitudinal surveys. We will also discuss tobacco use and tobacco control. In addition, this section will provide the background and demonstrate the significance of the problem to be solved. In Section III, we will introduce the mathematical model of smoking behavior, derive equations for estimating transitional probabilities, and apply the results to derive the probabilities from cross-sectional surveys. In Section IV, we will use the transitional probabilities to predict future smoking behavior. Such predictions will help us to evaluate various tobacco control programs. In Section V, we calculate various sensitivity functions from the model that we established. We emphasize that although this paper mainly focus on smoking behavior and tobacco control, the method established in this paper can be extended to describe other substance use behaviors. Indeed, the method of deriving transitional probabilities from cross-sectional survey data has a wide range of applications.

II. BACKGROUNDS AND MOTIVATIONS

In general, a longitudinal study or survey is a study that involves observations of the same subjects over long periods of time. Typically, longitudinal studies are used to study the progression of a behavior (such as smoking behavior) across the life span. The reason for this is that longitudinal studies obtain data at the individual level by tracking the same individuals, and the data can be used to determine the transitional dynamics and probabilities. There are different types of longitudinal studies. In this paper, we consider cohort studies. Such studies sample a cohort, defined as a group experiencing same event (birth, in our case) in a given time period, and make observations of the sampled subjects at intervals through time.

Cross-sectional studies or surveys, on the other hand, involve observation of random samples of a population at one point in time without follow-ups. They can be thought of as providing a "snapshot" of the frequency and characteristics of a behavior in a population at a particular point in time. For example, in a cross-sectional survey, a specific group is looked at to find predictor variables and outcome variables, such as smoking and lung cancer. Association analysis will then be conducted to establish the relationship between the predictor variables and the outcome variable. Also, cross-sectional data from health surveys are often used to describe health status of a study population, such as the prevalence rate of a disease (e.g., cancer) or a health risk behavior (e.g., smoking).

To study smoking behavior and to evaluate tobacco control programs, many surveys have been conducted in past decades. The majority of these surveys are cross-sectional in nature. In fact, a huge body of data on tobacco control has been accumulated. These data are often used to compute prevalence rates of smoking behavior (e.g., percentage of lifetime smoking, current smoking, or addicted smoking). A few studies examine the progression from nonsmokers to smokers using self-reported age of smoking onset from cross-sectional data [19–22]. But prevalence rates and probability of smoking onset alone are not enough, data on smoking

behavior progression must be added in order to increase the efficiency of tobacco control and to evaluate program effectiveness, because smoking behavior is an integrated, dynamic, and progressive process. With smoking progression data, it is possible to assess: a) the impact of a tobacco control effort (e.g., tobacco taxation, legal restrictions, school-based programs) on different steps of smoking behavior progression (e.g., from never-smokers to smokers or from regular smokers to quitters); b) the effect of changes in different steps of smoking progression (e.g. increasing quitter or reducing experimenters) in reducing the total number of smokers; and c) the amount of changes needed in different steps of smoking behavior progression in order to achieve a pre-determined tobacco control objective.

To provide data on smoking behavior progression for advanced tobacco control planning and program effect evaluation, a logic approach is to collect longitudinal data. Longitudinal data are often used in tobacco research to characterize smoking behaviors. For example, to characterize the history and trajectories of smoking behavior progression, longitudinal data have been collected in [46,16–18] to determine the risk factors that are associated with smoking behavior progression. For example, the California Tobacco Control Program conducted the California 1993–1996 Teen Longitudinal Survey. Also, the National Longitudinal Survey of Youth 1997 has collected longitudinal data in great details on smoking behavior progression and risk factors from subjects born in the years 1980–1984. The Monitoring the Future Studies has a biannual longitudinal data collection portion from senior high school student sample since 1976 to measure smoking behavior progression among youth. From longitudinal data, information can be directly derived to measure changes in smoking behavior according to progression stage and to determine the transitional probabilities that characterize the dynamics of smoking behavior progression in a population. Besides those examples, longitudinal data are not routinely collected in tobacco control practice.

Compared to a cross-sectional survey, a longitudinal survey is more time-consuming and difficult to perform because of the following reasons. (1) To follow up with study participants through a longitudinal survey is technically demanding even for professional tobacco researchers. In longitudinal surveys, the same individuals who participated in the baseline survey must be followed up at the subsequent times for data collection; consequently, strict and complicated procedures must be set up for correctly tracking the participants at the follow-ups while ensuring the confidentiality of the participants and the validity of the survey data. Consequently, compared to a cross-sectional survey, conducting a longitudinal survey needs more resources and personnel, especially personnel with advanced training and adequate practice. (2) It is more time-consuming to collect longitudinal data than to collect cross-sectional data. At least two waves of data collections are needed for a longitudinal survey to measure smoking behavior progression. It will take longer time to obtain information from a multi-wave longitudinal survey than from a one-wave cross-sectional survey for tobacco control planning and program effect evaluation.

There are also significant limitations to longitudinal data if they are used for tobacco control planning and program effect evaluation because of the following reasons. (1) Selection biases due to attrition: attrition or loss of follow-up is a common and significant concern with survey data collected through a longitudinal survey. Data from tobacco research indicate that participants who missed the follow-up are more likely to be smokers. This selective attrition will threaten the validity of longitudinal data. (2) Inaccuracy of survey time: for an ideal longitudinal survey, each wave of data collection should be completed at one time point (e.g., January 1, 2005 for wave 1 and January 1, 2006 for wave 2). However, a tobacco control program usually involves a population with large numbers of participants. Collecting data from such large samples can not be completed within one or two days, resulting in time errors in measuring smoking behavior progression even with advanced methodologies. For example, a participant may be surveyed once on January 1, 2005 and then again on March 1, 2006, instead

of January 1, 2006. This will cause a time error. (3) Hawthorne (survey) effect: repeatedly asking the same subjects the same questions regarding smoking behavior over time may result in biased data. (4) Recall biases: to obtain data on behavior dynamics, a longitudinal survey may ask each participant to recall in great details his or her smoking behavior in the past; this may result in erroneous data due to memory loss. (5) Age range of the subjects in a longitudinal sample shifts up as the subjects are followed up over time, affecting the use of such data in tobacco control practice [24–27,35,40,42].

Compared to a longitudinal survey, it will be easier and more cost-effective to conduct a cross-sectional survey in tobacco control practice. Collecting cross-sectional survey can be completed within a short period of time. The procedure to collect cross-sectional data is relatively simpler than to collect longitudinal data. Unlike longitudinal surveys, data from a cross-sectional survey can be analyzed without waiting for data from another wave. Cross-sectional data are generally less error-prone because items used in a cross-sectional survey often target the most recent events of tobacco use, such as smoking in the past 7 days and past 30 days; or events that are proven to be more accurately encoded in memory for recall, such as whether ever used a tobacco product, age when tried a tobacco product the first time or the last time in life. The validity of recalled data on these behaviors has been well documented [24–27]. In addition, data from cross-sectional survey are free from Hawthorne effect because each wave of the survey is conducted over different samples. Typical examples of cross-sectional surveys with data on tobacco smoking include National Survey on Drug Use and Health (NSDUH), Youth Risk Behavior Survey (YRBS), and Behavioral Risk Factor Survey (BRFS).

Although there are many advantages, cross-sectional data have never been used to assess smoking behavior progression because in any cross-sectional survey, no participants are followed up at next wave to collect the related data. However, considering one wave of cross-sectional survey as a snap shot of the smoking behavior dynamics in a population, our analysis indicates that data does contain information to assess smoking behavior progression. It would be ideal if a method is available to extract such information from cross-sectional data to measure the smoking behavior progression. This is indeed possible because at the aggregated level, data from a cross-sectional survey with a sample of subjects in multiple age groups is analogous to the data from a longitudinal survey that follows a sample of a birth cohort (born in one year) for multiple years. Hence information on smoking behavior progression at the aggregated level can be derived from cross-sectional data if appropriate methods are used.

III. PROBABILISTIC DISCRETE EVENT SYSTEMS

Discrete event system framework [31–34,38,41] provides a nice way to model smoking behavior progression using cross-sectional data. By incorporating probabilities into conventional discrete event system, we can derive a probabilistic discrete event system (PDES) for smoking behavior as

$$G=(Q, \Sigma, \delta, q_0).$$

The PDES model is illustrated in Figure 1. In the model, Q is the set of discrete states. In the smoking behavior model of Figure 1, $Q = \{NS, EX, SS, RS, QU\}$. Σ is the set of events. In Figure 1, $\Sigma = \{\sigma_1, \sigma_2, \dots, \sigma_{11}\}$, where each σ_i is an event describing the transition in smoking behavior. For example σ_2 is the event of starting smoking. $\delta : Q \times \Sigma \rightarrow Q$ is the transitional function describing what event can occur at which state and the resulting new states. For example, in Figure 1, $\delta(NS, \sigma_2) = EX$. q_0 is the initial state. For the smoking behavior model in Figure 1, $q_0 = NS$. With slight abuse of notation, we also use q to denote the probability of the

system being at state q and use σ_1 to denote the probability of σ_1 occurring. Therefore, NS also denotes the probability of being a never-smoker and σ_2 also denotes the probability of starting smoking. If it is important to specify the age, then we will use a to denote age. For example, $\sigma_2(a)$ denotes the event or the probability of starting smoking at age a .

From the PDES model of the smoking behavior shown in Figure 1, we can obtain the following equations.

$$\begin{aligned} NS(a+1) &= NS(a) - NS(a)\sigma_2(a) \\ EX(a+1) &= EX(a) + NS(a)\sigma_2(a) + SS(a)\sigma_5(a) - EX(a)\sigma_4(a) - EX(a)\sigma_7(a) \\ SS(a+1) &= SS(a) + EX(a)\sigma_4(a) - SS(a)\sigma_5(a) \\ RS(a+1) &= RS(a) + EX(a)\sigma_7(a) + QU(a)\sigma_{10}(a) - RS(a)\sigma_9(a) \\ QU(a+1) &= QU(a) + RS(a)\sigma_9(a) - QU(a)\sigma_{10}(a) \end{aligned}$$

For example, the first equation above states that the percentage (or probability) of people who are never-smoker at age $a + 1$ is equal to the percentage of people who are never-smoker at age a subtract the percentage of people who are never-smoker at age a times the percentage of never-smokers who start smoking at age a . The other equations can be interpreted similarly. In addition, since probabilities must sum up to 1, we have the following obvious equations (with reference to Figure 1).

$$\begin{aligned} \sigma_1(a) + \sigma_2(a) &= 1 \\ \sigma_3(a) + \sigma_4(a) + \sigma_7(a) &= 1 \\ \sigma_5(a) + \sigma_6(a) &= 1 \\ \sigma_8(a) + \sigma_9(a) &= 1 \\ \sigma_{10}(a) + \sigma_{11}(a) &= 1. \end{aligned}$$

Write the above 10 equations in matrix form $A\sigma = B$ as Equation 1. Then it can be checked that the rank of A is 9: $rank(A) = 9$. Therefore, only 9 equations are independent.

$$\begin{bmatrix} 0 & -NS(a) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & NS(a) & 0 & -EX(a) & SS(a) & 0 & -EX(a) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & EX(a) & -SS(a) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & EX(a) & 0 & -RS(a) & QU(a) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & RS(a) & -QU(a) & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \sigma_1(a) \\ \sigma_2(a) \\ \sigma_3(a) \\ \sigma_4(a) \\ \sigma_5(a) \\ \sigma_6(a) \\ \sigma_7(a) \\ \sigma_8(a) \\ \sigma_9(a) \\ \sigma_{10}(a) \\ \sigma_{11}(a) \end{bmatrix} = \begin{bmatrix} NS(a+1) - NS(a) \\ EX(a+1) - EX(a) \\ SS(a+1) - SS(a) \\ RS(a+1) - RS(a) \\ QU(a+1) - QU(a) \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

(Equation

1)

Since there are 11 transitional probabilities, $\sigma_1(a)$, $\sigma_2(a)$, ..., $\sigma_{11}(a)$, to be solved, we need 2 more independent equations. These equations can be obtained as follows. In the NSDUH survey, subjects were asked the last time when they smoked. Using the data, we estimate the portion of the self stoppers who stopped smoke 24 months ago (and hence are self stoppers a year ago) or "old self stoppers". Denote the old self stoppers as \overline{SS} . Then we have another equation:

$$\overline{SS}(a+1)=SS(a)\sigma_6(a).$$

Similarly, denote the old quitter as \overline{QU} and we have one more equation:

$$\overline{QU}(a+1)=QU(a)\sigma_{11}(a).$$

The above 2 equations, plus the 9 independent equations, will allow us to solve for all transitional probabilities as shown in Equation 2.

$$\begin{bmatrix} \sigma_1(a) \\ \sigma_2(a) \\ \sigma_3(a) \\ \sigma_4(a) \\ \sigma_5(a) \\ \sigma_6(a) \\ \sigma_7(a) \\ \sigma_8(a) \\ \sigma_9(a) \\ \sigma_{10}(a) \\ \sigma_{11}(a) \end{bmatrix} = \begin{bmatrix} 0 & -NS(a) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & NS(a) & 0 & -EX(a) & SS(a) & 0 & -EX(a) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & EX(a) & -SS(a) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & EX(a) & 0 & -RS(a) & QU(a) & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & SS(a) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & QU(a) \end{bmatrix}^{-1} \begin{bmatrix} NS(a+1) - NS(a) \\ EX(a+1) - EX(a) \\ SS(a+1) - SS(a) \\ RS(a+1) - RS(a) \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ \overline{SS}(a+1) \\ \overline{QU}(a+1) \end{bmatrix}$$

(Equation 2)

Let us now use the data from 2000 NSDUH to calculate these transitional probabilities. The data from age 15 to 21 are shown in Table I. We use the percentages of people in various states as the state probabilities.

We calculate the transitional probabilities from age 15 to 21 for the data in Table I using Equation 2. The results are shown in Table II.

These transitional probabilities describe the dynamics of smoking behavior among US adolescents and young adults in 2000. They can be used to evaluate and hence improve the tobacco control programs as to be discussed in the next two sections.

IV. PREDICTING SMOKING BEHAVIOR FOR TOBACCO CONTROL PROGRAMS

The transitional probabilities provide information on the likelihood that a person may start smoking, progress toward a regular smoker, quit smoking, etc. These transitional probabilities are influenced by the environment which the person is in. Various tobacco control programs, such as tobacco taxation, restriction of smoking in public places, restriction of tobacco sales to minors, school-based programs, and media campaigns, are intended to change the environment and hence the transitional probabilities. Different tobacco control programs have different impacts on the transitional probabilities. For example, restrictions of tobacco sales to minors and school-based programs have greater impact on $\sigma_2(a)$ than on other transitional probabilities. The goal of tobacco control programs is to reduce smoking among adolescents and adults. In terms of PDES, the goal is to reduce the (state) probability RS . To qualitatively assess the impact of a tobacco control program to RS , we need to do some prediction on how

transitional probabilities $\sigma_i(a)$ impact on probability RS . This can be done as follows. Suppose $\sigma_i(a)$ is changed to $\sigma'_i(a)$. Denote the new transition matrix as

$$\Pi'(a) = \begin{bmatrix} \sigma'_1(a) & 0 & 0 & 0 & 0 \\ \sigma'_2(a) & \sigma'_3(a) & 0 & \sigma'_5(a) & 0 \\ 0 & \sigma'_7(a) & \sigma'_8(a) & 0 & \sigma'_{10}(a) \\ 0 & \sigma'_4(a) & 0 & \sigma'_6(a) & 0 \\ 0 & 0 & \sigma'_9(a) & 0 & \sigma'_{11}(a) \end{bmatrix}.$$

Let the state probabilities at ages a and $a + 1$ under the new transitional probabilities $\pi'(a)$ be denoted by

$$Q'(a) = \begin{bmatrix} NS'(a) \\ EX'(a) \\ RS'(a) \\ SS'(a) \\ QU'(a) \end{bmatrix} \quad Q'(a+1) = \begin{bmatrix} NS'(a+1) \\ EX'(a+1) \\ RS'(a+1) \\ SS'(a+1) \\ QU'(a+1) \end{bmatrix}$$

respectively. Then we can predicate the state probabilities of smoking behavior at different ages as follows.

$$\begin{aligned} Q'(a+1) &= \Pi'(a)Q'(a) \\ Q'(a+2) &= \Pi'(a+1)\Pi'(a)Q'(a) \\ \dots \\ Q'(a+k) &= \Pi'(a+k-1) \dots \Pi'(a+1)\Pi'(a)Q'(a) \end{aligned}$$

For example, if a tobacco control program can decrease the probability of σ_2 by 10%, then we can recalculate the transitional probabilities from age 15 to 21 and predict the state probabilities of smoking behavior accordingly. The predicted results are shown in Table III. By comparing Table III with Table I, we can estimate the effects of the tobacco control program.

V. SENSITIVITY ANALYSIS

To further understand the impact of transitional probabilities $\sigma_i(a)$ at age a on state probabilities $q(b)$ at age b (q can be any state), we would like to find the sensitivities from $\sigma_i(a)$ to $q(b)$,

which is represented by $\frac{dq(b)}{d\sigma_i(a)}$. Derivative $\frac{dq(b)}{d\sigma_i(a)}$ tells us how the change in $\sigma_i(a)$ at age a will affect $q(b)$ at age $b > a$. To this end, let us write the smoking behavior equation as

$$Q(b) = \Pi(b-1) \dots \Pi(a+1)\Pi(a)Q(a).$$

Then

$$\frac{dQ(b)}{d\sigma_i(a)} = \Pi(b-1) \dots \Pi(a+1) \frac{d\Pi(a)}{d\sigma_i(a)} Q(a).$$

Since

$$\Pi(a) = \begin{bmatrix} \sigma_1(a) & 0 & 0 & 0 & 0 \\ \sigma_2(a) & \sigma_3(a) & 0 & \sigma_5(a) & 0 \\ 0 & \sigma_7(a) & \sigma_8(a) & 0 & \sigma_{10}(a) \\ 0 & \sigma_4(a) & 0 & \sigma_6(a) & 0 \\ 0 & 0 & \sigma_9(a) & 0 & \sigma_{11}(a) \end{bmatrix}$$

and some $\sigma_i(a)$ depends on some other $\sigma_j(a)$, $\frac{d\Pi(a)}{d\sigma_i(a)}$ is calculated by first replacing dependency of $\sigma_i(a)$ and then taking derivative. For example, for transition $\sigma_2(a)$,

$$\frac{d\Pi(a)}{d\sigma_2(a)} = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Using the data in Table I and Table II, we can calculate sensitivities $\frac{dq(b)}{d\sigma_i(a)}$. For example, let

$a=15$ and $b=17, 19, 21$, we can calculate $\frac{dQ(b)}{d\sigma_2(a)}$ as follows.

$$\frac{dQ(17)}{d\sigma_2(15)} = \begin{bmatrix} -56.0120 \\ 15.2760 \\ 24.1870 \\ 17.1855 \\ 0 \end{bmatrix}, \quad \frac{dQ(19)}{d\sigma_2(15)} = \begin{bmatrix} -42.3899 \\ -1.2627 \\ 30.2190 \\ 12.3877 \\ 1.6824 \end{bmatrix}, \quad \frac{dQ(21)}{d\sigma_2(15)} = \begin{bmatrix} -36.2179 \\ -1.2546 \\ 28.8527 \\ 6.8274 \\ 2.4290 \end{bmatrix}.$$

To see the percentage reduction of $\sigma_2(a)$ (less initiation of smoking) on the percentage of never-smokers $NS(a)$, we calculate the following sensitivity function from $\sigma_2(a)$ to $NS(b)$

$$\frac{\frac{dNS(b)}{NS(b)}}{-\frac{d\sigma_2(a)}{\sigma_2(a)}} = -\frac{dNS(b)}{d\sigma_2(a)} \frac{\sigma_2(a)}{NS(b)}$$

Using the data in Table I and Table II, the results are as follows.

$$\frac{\frac{dNS(17)}{NS(17)}}{-\frac{d\sigma_2(15)}{\sigma_2(15)}} = 20.28\%, \quad \frac{\frac{dNS(19)}{NS(19)}}{-\frac{d\sigma_2(15)}{\sigma_2(15)}} = 20.27\%, \quad \frac{\frac{dNS(21)}{NS(21)}}{-\frac{d\sigma_2(15)}{\sigma_2(15)}} = 20.29\%.$$

Similarly, we can calculate the sensitivity function from $\sigma_2(a)$ to $RS(a)$ as follows.

$$\frac{\frac{dRS(17)}{RS(17)}}{-\frac{d\sigma_2(15)}{\sigma_2(15)}} = -24.19\%, \quad \frac{\frac{dRS(19)}{RS(19)}}{-\frac{d\sigma_2(15)}{\sigma_2(15)}} = -32.33\%, \quad \frac{\frac{dRS(21)}{RS(21)}}{-\frac{d\sigma_2(15)}{\sigma_2(15)}} = -28.52\%.$$

Clearly from these sensitivities, we can conclude that reducing $\sigma_2(a)$ (less initiation of smoking) will substantially increase the number of never-smokers and substantially decrease the number of regular smokers, although the impacts are different at different age.

VI. CONCLUSION

In this paper, we reported a new approach to investigate longitudinal smoking behavior progression using cross-sectional data. We derived a smoking behavior model based on PDES. Using this model, we then estimated the transitional probabilities from the survey data of 2000 NSDUH. There are several important applications of this model. Using the estimated transitional probabilities, we can predict the smoking behavior with respect to the changes in transitional probabilities for certain age groups. Knowing the effects of tobacco control programs to the transitional probabilities, we can then evaluate various tobacco control programs and hence provide assistance to the policy makers in their decision making. We also derived sensitivity functions of various transitional probabilities to the state probabilities. These functions can serve as analytical tools for comparing the effects of different transitional probabilities on the final outcome of tobacco control. Overall, the paper shows that transitional probabilities can be estimated from cross-sectional survey data and hence to describe the dynamics of behavior progression systems. The establishment of this method will open a new direction in behavior science research beyond tobacco smoking.

Acknowledgments

This research is supported in part by National Science Foundation under grants ECS-0624828 and ECS-0823865, and by National Institute of Health under grant 1R01DA022730-01A2. We would like to thank George Yin for several inspiring discussions on the subject.

REFERENCES

1. CDC. U.S. DHHS, CDC National Center for Chronic Disease Prevention and Health Promotion. Atlanta, GA: Office on Smoking and Health; 1999. Best Practices for Comprehensive Tobacco Control Programs - August 1999.
2. CDC. US DHHS Office of Disease Prevention and Health Promotion. Rockville, MD: 2000. Healthy People 2010.
3. CDC. Introduction to Program Evaluation for Comprehensive Tobacco Control Programs. Atlanta, GA: U.S. DHHS; 2001.
4. CDC. Global Tobacco Control Program. Atlanta, GA: CDC Office on Smoking and Health; 2004.
5. DHHS, U. US DHHS, Office on Smoking and Health, Centers for Disease Control and Prevention. Atlanta, GA: 1992. Reducing the Health Consequences of Smoking: 25 Years of Progress - A Report of the Surgeon General.
6. US DHHS. US DHHS, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion. Atlanta, GA: Office on Smoking and Health; 1994. Preventing Tobacco Use among Young People - A Report of the Surgeon General.
7. US DHHS. US DHHS, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion. Atlanta, GA: Office on Smoking and Health; 2000. Reducing Tobacco Use: A Report of the Surgeon General.
8. US DHHS. US DHHS, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion. Atlanta, GA: Office on Smoking and Health; 2004. The Health Consequences of Smoking - A Report of the Surgeon General.
9. US Department of Agriculture. U.S. Department of Agriculture. Springfield, VA: Economic Research Services; 2003. Tobacco Outlook.
10. Bauman KE, Ennett ST. Validity of adolescent self-reports of cigarette smoking. *Am J Public Health* 1998;88(2):309–310. [PubMed: 9491034]
11. Bernstein DM, Whittlesea BW, Loftus EF. Increasing confidence in remote autobiographical memory and general knowledge: extensions of the revelation effect. *Mem Cognit* 2002;30(3):432–438.
12. Bhowal P, Sarkar D, Mukhopadhyay S, Basu A. Fault diagnosis in discrete time hybrid systems - A case study. *Information Sciences* 2007;177(5):1290–1308.

13. Brauer LH, Hatsukami D, Hanson K, Shiffman S. Smoking topography in tobacco chippers and dependent smokers. *Addict Behav* 1996;21(2):233–238. [PubMed: 8730526]
14. California Department of Health Services. Cancer Prevention and Control Unit. La Jolla, CA: University of California, San Diego; 1999. California Tobacco Survey: 1993–1996 Teen Longitudinal Survey.
15. Chaloupka FJ, Wechsler H. Price, tobacco control policies and smoking among young adults. *J Health Econ* 1997;16(3):359–373. [PubMed: 10169306]
16. Chassin L, Presson CC, Sherman SJ, Edwards DA. The natural history of cigarette smoking: predicting young-adult smoking outcomes from adolescent smoking patterns. *Health Psychol* 1990;9(6):701–716. [PubMed: 2286181]
17. Chassin L, Presson CC, Pitts SC, Sherman SJ. The natural history of cigarette smoking from adolescence to adulthood in a midwestern community sample: multiple trajectories and their psychosocial correlates. *Health Psychol* 2000;19(3):223–231. [PubMed: 10868766]
18. Chen K, Kandel DB. The natural history of drug use from adolescence to the mid-thirties in a general population sample. *Am J Public Health* 1995;85(1):41–47. [PubMed: 7832260]
19. Chen X, Li G, Unger JB, Liu X, Johnson CA. Secular trends in adolescent never smoking from 1990 to 1999 in California: an age-period-cohort analysis. *Am J Public Health* 2003;93(12):2099–2104. [PubMed: 14652341]
20. Chen X, Stanton B, Shankaran S, Li X. Age of smoking onset as a predictor of smoking cessation during pregnancy. *American Journal of Health Behavior*. 2006 In press.
21. Chen X, Li X, Stanton B, Mao R, Sun Z, Zhang H, Qu M, Wang J, Thomas R. Patterns of cigarette smoking among students from 19 colleges and universities in Jiangsu Province, China: a latent class analysis. *Drug Alcohol Depend* 2004;76(2):153–163. [PubMed: 15488339]
22. Chen X, Li Y, Unger JB, Gong J, Johnson CA, Guo Q. Hazard of smoking initiation by age among adolescents in Wuhan, China. *Prev Med* 2001;32(5):437–445.
23. Ignjatović J, Ćirić M, Stojan Bogdanović S. Determinization of fuzzy automata with membership values in complete residuated lattices. *Information Sciences* 2008;178(1):164–180.
24. Janson H. Longitudinal patterns of tobacco smoking from childhood to middle age. *Addict Behav* 1999;24(2):239–249. [PubMed: 10336105]
25. Johnson TP, Mott JA. The reliability of self-reported age of onset of tobacco, alcohol and illicit drug use. *Addiction* 2001;96(8):1187–1198. [PubMed: 11487424]
26. Kaplan RM, Ake CF, Emery SL, Navarro AM. Simulated effect of tobacco tax variation on population health in California. *Am J Public Health* 2001;91(2):239–244. [PubMed: 11211632]
27. Korkeila K, Suominen S, Ahvenainen J, Ojanlatva A, Rautava P, Helenius H, Koskenvuo M. Non-response and related factors in a nation-wide health survey. *Eur J Epidemiol* 2001;17(11):991–999. [PubMed: 12380710]
28. Kilic E. Diagnosability of fuzzy discrete event systems. *Information Sciences* 2008;178(3):858–870.
29. Lei H, Li Y. Minimization of states in automata theory based on finite lattice-ordered monoid. *Information Sciences* 2007;177(6):1413–1421.
30. Liu J, Li Y. The relationship of controllability between classical and fuzzy discrete-event Systems. *Information Sciences* 2008;178(2):4142–4151.
31. Lin F, Wonham WM. “On observability of discrete-event systems, ”. *Information Sciences* 1988;44(3):173–198.
32. F Lin F, Wonham WM. “Decentralized supervisory control of discrete-event systems, ”. *Information Sciences* 1988;44(3):199–224.
33. Lin F, Ying H, MacArthur RD, Cohn JA, Barth -Jones DC, Ye H, Crane LR. “Theory for a control architecture of fuzzy discrete event systems for decision making, ”. *Information Science* 2007;177:3749–3763.
34. Mortazavian H, Lin F. “Decentralized supervisory control of discrete event systems with nonhomogeneous control structure, ”. *Information Sciences* 1993;68(3):233–246. 1993.
35. Murray M, Swan AV, Kiryluk S, Clarke GC. The Hawthorne effect in the measurement of adolescent smoking. *J Epidemiol Community Health* 1988;42(3):304–306. [PubMed: 3251014]

36. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. *J Consult Clin Psychol* 1983;51(3):390–395. [PubMed: 6863699]
37. Ramadge PJ, Wonham WM. Supervisory control of a class of discrete event processes. *SIAM Journal of Control and Optimization* 1987;25(1):206–230.
38. Ramadge PJ, Wonham WM. The control of discrete event systems. *Proceedings of the IEEE* 1989;77(1):81–98.
39. SAMHA. Overview of Key Findings, 2004. Rockville, MD: U.S. DHHS; 2005. Monitoring the Future National Results on Adolescent Drug Use.
40. Singer, JD.; Willett, JB. *Applied Longitudinal Data Analysis*. New York: Oxford University Press; 2003.
41. Shu S, Lin F, Ying H, Chen X. “State estimation and detectability of probabilistic discrete event systems,”. *Automatica* 2008;44(12):3054–3060. [PubMed: 19956775]
42. Stanton WR, McClelland M, Elwood C, Ferry D, Silva PA. Prevalence, reliability and bias of adolescents' reports of smoking and quitting. *Addiction* 1996;91(11):1705–1714. [PubMed: 8972928]
43. Starr, G.; Rogers, T.; Schooley, M.; Porter, S.; Wiesen, E.; Jasmison, N. Centers for Disease Control and Prevention. Atlanta, GA: 2005. Key Outcome Indicators for Evaluating Comprehensive Tobacco Control Programs.
44. van der Weide, ThP; van Bommel, P. Measuring the incremental information value of documents. *Information Sciences* 2006;176(2):91–119.
45. Weinberger DA, Tublin SK, Ford ME, Feldman SS. Preadolescents' social-emotional adjustment and selective attrition in family research. *Child Dev* 1990;61(5):1374–1386. [PubMed: 2245731]
46. White HR, Pandina RJ, Chen PH. Developmental trajectories of cigarette use from early adolescence into young adulthood. *Drug Alcohol Depend* 2002;65(2):167–178. [PubMed: 11772478]

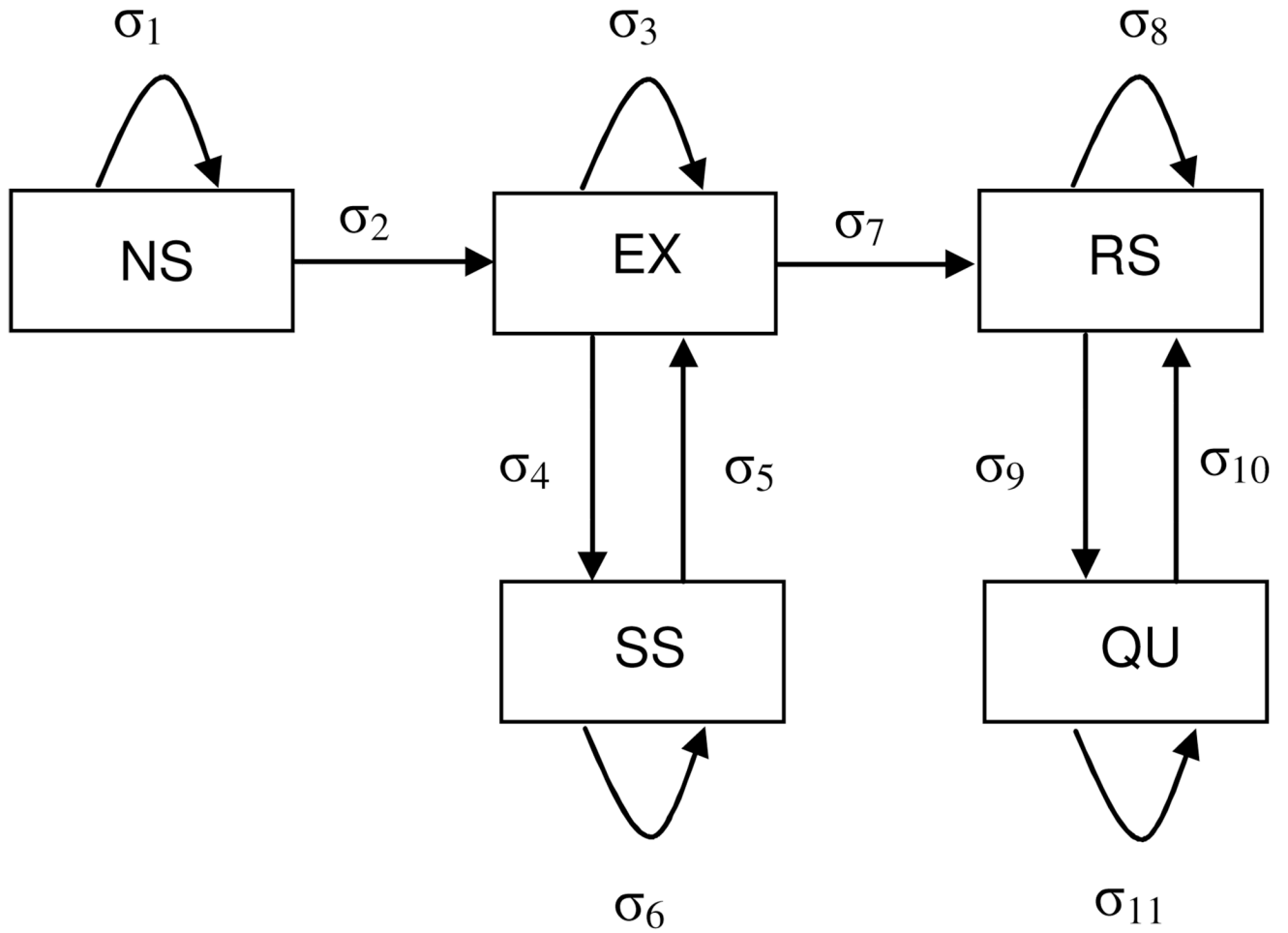


Fig. 1. Probabilistic discrete event system model of the smoking behavior. States are: NS - nerve smoker, EX - experimenter, SS - self stopper, RS - regular smoker, and QU - quitter. σ_i are events and corresponding probabilities of transitions among states. NS is the initial state.

Table 1

Smoking data from 2000 National Survey on Drug Use and Health (NSDUH) The data show the percentages of people in different states.

Age	NS	EX	SS	RS	QU	SS	$\bar{Q}U$
15	63.65	12.81	14.74	7.84	0.66	8.61	0.42
16	53.10	15.57	17.69	12.45	0.88	12.36	0.40
17	46.95	16.56	17.00	17.99	1.18	12.83	0.54
18	41.20	16.11	16.40	24.46	1.64	11.24	0.87
19	35.55	15.89	15.89	30.50	2.08	11.83	1.34
20	31.75	15.09	16.05	34.69	2.36	12.29	1.51
21	30.35	13.69	17.20	35.77	2.94	13.05	1.73

Table II

Transitional probabilities derived from 2000 NSDUH

Age	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	σ_7	σ_8	σ_9	σ_{10}	σ_{11}
15	0.83	0.17	0.21	0.42	0.16	0.84	0.38	0.94	0.06	0.38	0.62
16	0.88	0.12	0.36	0.27	0.27	0.73	0.38	0.95	0.05	0.38	0.62
17	0.88	0.12	0.27	0.31	0.34	0.66	0.42	0.96	0.04	0.26	0.74
18	0.86	0.14	0.35	0.25	0.28	0.72	0.40	0.97	0.03	0.18	0.82
19	0.89	0.11	0.48	0.24	0.23	0.77	0.28	0.97	0.03	0.28	0.72
20	0.96	0.04	0.62	0.27	0.19	0.81	0.11	0.97	0.03	0.27	0.73

Table III
 Prediction of smoking behavior if some tobacco control program can decrease the probability of σ_2 (α) by 10%

Age	NS	EX	SS	RS	QU
15	63.65	12.81	14.74	7.84	0.66
16	53.91	14.79	17.76	12.49	0.88
17	48.09	15.94	16.96	17.82	1.17
18	42.90	15.26	16.13	24.10	1.58
19	37.49	15.26	15.43	29.77	2.02
20	33.78	14.59	15.55	33.72	2.35
21	32.56	13.21	16.53	34.94	2.72