Systemize the Probabilistic Discrete Event Systems with Moore-penrose Generalized-inverse Matrix Theory for Cross-sectional Behavioral Data

Ding-Geng (Din) Chen[1,2]*, Xinguang Chen[3], Feng Lin[4], Y.L. Lio[5] and Harriet Kitzman[1]

[1]Center for Research, School of Nursing, USA
[2]Department of Biostatistics and Computational Biology, University of Rochester Medical Center, Rochester, USA
[3]Department of Epidemiology, University of Florida, Gainsville, FL, USA
[4]Department of Electrical and Computer Engineering, Wayne State University, Detroit, MI, USA
[5]Department of Mathematical Sciences, University of South Dakota, SD, USA

Abstract

Moore-Penrose (M-P) generalized inverse matrix theory provides a powerful approach to solve an admissible linear-equation system when the inverse of the coefficient matrix does not exist. M-P matrix theory has been used in different areas to solve challenging research questions, including operations research, signal process, and system controls. In this study, we report our work to systemize a probability discrete event systems (PDES) modeling in characterizing the progression of health risk behaviors. A novel PDES model was devised by Lin and Chen to extract and investigate longitudinal properties of smoking multi-stage behavioral progression with cross-sectional survey data. Despite its success, this PDES model requires extra exogenous equations for the model to be solvable and practically implementable. However, exogenous equations are often difficult if not impossible to obtain. Even if the additional exogenous equations are derived, the data used to generate the equations are often error-prone. By applying the M-P theory, our research demonstrates that Lin and Chen's PDES model can be solved without using exogenous equations. For practical application, we demonstrate the M-P approach using the open-source R software with real data from 2000 National Survey of Drug Use and Health. The removal of extra data facilitate researchers to use the novel PDES method in examining human behaviors, particularly, health related behaviors for disease prevention and health promotion. Successful application of the M-P matrix theory in solving the PDES model suggests potentials of this method in system modeling to solve challenge problems for other medical and health related research.

Background

The Moore-Penrose (M-P) generalized inverse matrix theory [1,2] provides a powerful tool to solve a liner equation system that cannot be solved by using the inverse of the coefficient matrix. Although M-P matrix theory has been used to solve challenging problems in operations research, signal process, system controls and various other fields [3-7], to date this method has not been used in health and human behavior research. In this study, we report our work to solve a probability discrete event system-based modeling characterizing cigarette smoking behavior among an adolescent population in the United States.

To extract and model the longitudinal properties of multi-stage behavioral system, such as cigarette smoking with cross-sectional survey data, Chen et al. [8-10] developed the probability discrete event systems(PDES) modeling approach. In this approach, the continuous development process of a behavior (such as, cigarette smoking, disease progression) is first conceptualized as a PDES with multiple states. These states describe the multiple stages of logic behavioral progression with the transition paths linking one state (stage) to another [8-10]. This model has been successfully used in describing the dynamics of cigarette smoking behavior [8,9] and the responses to smoking prevention intervention among adolescents in the United States [11]. Despite the success, the established PDES method has a limitation: the model cannot be determined without extra exogenous equations. Furthermore, such exogenous equations are often impractical to obtain and even if an equation is derived, the data supporting the construction of the equation may be error prone.

To overcome the limitation of the PDES modeling method, we proposed the use of M-P inverse matrix method that can solve the established PDES model without exogenous equation (s) to create a full-ranked coefficient matrix. The combined approach of the M-P inverse matrix theory with PDES (or "M-P Approach" for short) will increase the efficiency and utility of PDES modeling in investigating many dynamics of human behavior without fully observed data. To facilitate the use of the M-P Approach, an R program with examples and data are provided in Appendix A for interested readers to apply their own research data.

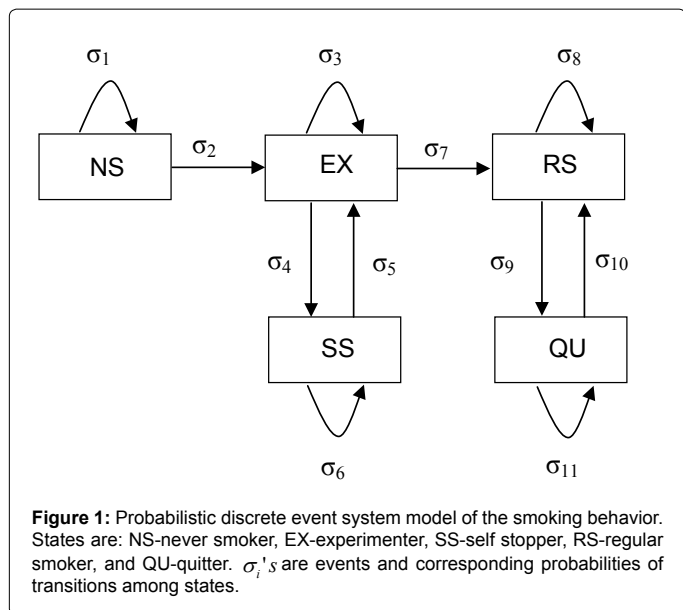## A Review of the PDES for Smoking Behavior

To be self-contained, we make use of the notations in Lin and Chen [10] in this paper to describe the PDES model. According to Lin and Chen [10], in estimating the transitional probability with cross-sectional survey data to model smoking multi-behavioral progression (Figure 1), five behavioral states are defined to construct a PDES:

Figure 1: Probabilistic discrete event system model of the smoking behavior. States are: NS-never smoker, EX-experimenter, SS-self stopper, RS-regular smoker, and QU-quitter. $\sigma_i's$ are events and corresponding probabilities of transitions among states.

- NS – never-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.

- QU – quitter, a regular smoker who stopped smoking for at least 12 months.

The smoking dynamics as shown in Figure 1 can be described using the PDES model:

$$G = (Q, \Sigma, \delta, q_0) \tag{1}$$

where Q is the set of discrete states. In this smoking behavior model of Figure 1. $Q=\{NS,EX,SS,RS,QU\}$ Let $\Sigma=\{\sigma_1, \sigma_2, ..., \sigma_{11}\}$ be the set of events. In Figure 1, $\Sigma=\{\sigma_1, \sigma_2, ..., \sigma_{11}\}$, where each σ_i is an event describing the transition among the multiple smoking behaviors. 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, δ $(NS, \sigma2)=EX$. 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_i to denote the probability of σ_i occurring. Therefore, NS also denotes the probability of being a never-smoker and σ_i 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.

Based on the defined PDES model shown in Figure 1, the following equationset can be defined conceptually:

$$NS (a+1) = NS(a)-NS(a)\,\sigma_2\,(a) \tag{2}$$

$$EX (a+1) = EX (a) + NS (a)\,\sigma_2\,(a) + SS (a)\,\sigma_5\,(a) - EX (a)\,\sigma_4(a) - EX(a)\,\sigma_7(a) \tag{3}$$

$$SS (a+1) = SS(a)+ EX (a)\,\sigma_4\,(a) - SS (a)\sigma_5(a) \tag{4}$$

$$RS (a+1) = RS(a)+ EX(a)\sigma_7(a)+ QU (a)\,\sigma_{10}(a) - RS(a)\,\sigma_9\,(a) \tag{5}$$

$$QU(a+1) = QU(a) + RS (a)\,\sigma_9(a) - QU (a)\,\sigma_{10}\,(a) \tag{6}$$

For example, Equation (2) states that the percentage of people who are never-smoker at age $a+1$ is equal to the percentage of people who are never-smoker at age a, subtracted from the percentage of people who are never-smoker at age a, times the percentage of never-smokers who start smoking at age a. Similar explanations can be done for the other equations. Furthermore, we have the following additional equations with respect to Figure 1.

$$\sigma_1\,(a) + \sigma_2\,(a)=1 \tag{7}$$

$$\sigma_3\,(a) + \sigma_4\,(a)+ \sigma_7\,(a)=1 \tag{8}$$

$$\sigma_5\,(a) + \sigma_6\,(a)=1 \tag{9}$$

$$\sigma_8\,(a) + \sigma_9\,(a)=1 \tag{10}$$

$$\sigma_{10}\,(a) + \sigma_{11}\,(a)=1 \qquad . \tag{11}$$

The above 10 equations from Equation (2) to Equation (11) can be casted into the matrix format:

$$\begin{bmatrix} 0 & -NS(a) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & NS(a) & 0 & -EX(a) & SS(a) & 0 & -EX(a) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & EX(a) & -SS(a) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & EX(a) & 0 & -(RSa) & QU(a) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & RS(a) & QU(a) & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \sigma_4 \\ \sigma_5 \\ \sigma_6 \\ \sigma_7 \\ \sigma_8 \\ \sigma_9 \\ \sigma_{10} \\ \sigma_{11} \end{bmatrix} \tag{12}$$

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

Equation (12) is denoted by $A\sigma=b$ where A is the coefficient matrix, σ the bolded is the solution vector and vector b denotes the right-side of Equation (12).

It can be shown that *rank (A)=9*. Therefore, among the 10 equations, only 9 are independent. However there are 11 transitional probabilities, $\sigma_1(a), \sigma_2(a), ..., \sigma_{11}(a)$ to be estimated. Therefore the PDES equation set (12) cannot be solved uniquely as indicated in Lin and Chen [10]. This condition will restrict the application of this novel approach in research and practice.

To solve this challenge, Lin and Chen [10] sought to derive two more independent equations by squeezing the survey data to define two additional progression stages (1) $\overline{SS}$, old self-stoppers (e.g., those who stopped smoking one year ago) and (2) $\overline{QU}$ old quitters (e.g., those who quit smoking one year ago). With data for these two newly defined smokers, two more independent equations $\overline{SS}(a+1)=SS(a)\sigma_6(a)$ and $\overline{QC}(a+1)=QC(a)\sigma_{11}(a)$ are derived to ensure the equation set (12)

has a definite solution. However, the introduction of the two types of smokers $\overline{SS}$ and $\overline{QC}$ may have also brought in more errors from the data because two newly defined smokers must be derived from recalled data one year longer than other data. If this is the case, errors introduced through these two newly defined smokers will affect the estimated transitional probabilities that are related to self-stoppers and quitters, including σ_3, σ_4, σ_5, σ_6, σ_{10}, and σ_{11} (Figure 1). When searching for methods that can help to solve Equation (12) without depending on the two additional equations, we found the generalized inverse matrix approach [1,12]. It is this "M-P Approach" that makes the impossible PDES work possible.

Generalized-inverse Matrix for PDES

In matrix theory, the generalized-inverse of a matrix A with dimension $m \times n$ (i.e. m rows with m equations and n columns with n variables) is defined as: $AA^-A = A$ where A^- is called the generalized-inverse of A. The purpose of introducing a generalized-inverse for any matrix is to have a general solution $\sigma = A^-b$ for any linear system $A\sigma = b$ (in corresponding to the PDES described in Equation 12) regardless of the existence of the inverse of coefficient matrix A. With this extension, if A is invertible, i.e. A^{-1} exists, the linear system $A\sigma = b$ is equivalent to the classical solution $\sigma = A^{-1}b$ as commonly known in any elementary linear algebra course. From the definition of the generalized inverse matrix, it can be seen that $A^- = A^{-1}$ if A is a full-rank square matrix. In this case, rank (A)=m=n. Obviously as described earlier, the matrix A for the PDES system (e.g., Equation 12) is not a full-rank matrix (i.e. rank(A) is less than m, n), in another word, the system is complete but the observed data to support solving the system is incomplete. Therefore a system without fully observed data like the PDES model cannot be solved using the classic matrix approach. With the introduction of the generalized-inverse matrix approach, we will show that for any matrix equation $A\sigma = b$, including the PDES described in Equation 12:

$\sigma = A^-b$ is a solution to $A\sigma = b$.

The general solution to the PDES matrix equation of $A\sigma = b$ can be expressed in $\sigma = A^-b + (I - A^-A)z$ where A^{-1} is any fixed generalized-inverse of A, while z represents an arbitrary vector. Therefore, the generalized-inverse A^{-1} is not unique which is equivalent to say that the PDES equation system (12) cannot be solved uniquely as indicated in Lin and Chen [10]. To practically solve this challenge, Lin and Chen [10] sought to derive two exogenous equations in order to solve for 11 parameters. However, the data used to construct those exogenous equations are hard to obtain and error-prone. Inspired by the general inverse matrix theory, particularly the work by Moore and Penrose, we introduced a mathematical approach to this problem: the M-P Approach. In his famous paper, Moore proposed three more conditions to the generalized-inverse A^- defined above. They are as follows:

$$AA^-A = A$$

The original definition of generalized-inverse matrix is to allow any admissible linear system $A\sigma = b$ to be solved easily by matrix representation regardless of the existence of the inverse of coefficient matrix. Extending the classical inverse matrix definition, $AA^{-1} = 1$ with the identity matrix I, which is equivalent to $AA^{-1}A = (AA^{-1})A = IA = A$, AA^{-1} is relaxed and no longer needs to be an identity matrix. With this extension, the only requirement is that AA^{-1} will map all column vectors of A to the same column vectors, respectively.

This added condition makes A^- a generalized reflexive inverse of A. Similar to the original definition of a generalized-inverse matrix, this added condition is to guarantee that the classical inverse matrix

definition of $A^{-1}A = I$ can still hold from this generalized-inverse so that $A^{-1}AA^{-1} = A^{-1}$ when the inverse exists. With this condition $A\,A$ does not need to be an identity matrix, but to map all column vectors of A^- to the same column vectors, respectively.

$$(AA^-)' = AA^-$$

The third condition addresses the transpose of AA^- to be itself. It indicates that AA^- is a Hermitian matrix. This is intuitively true that when A is invertible, $AA^- = AA^{-1} = I$ and the transpose of identity matrix I is itself

$$(A^-A)' = A^-A$$

The fourth condition is similar to the third condition. It indicates that A^-A is a Hermitian matrix with an intuitive explanation similar to the third condition.

Moore's extended definition did not receive any attention in the mathematics field for twenty years until Penrose [2] proved the uniqueness of Moore's definition. Since Penrose's work, this definition has been named as Moore-Penrose generalized-inverse and is typically denoted as A^+. The Moore-Penrose generalized inverse has several mathematical properties, and the most relevant one to PDES is that the solution of $\sigma = A^+b$ is unique (Appendix B.1) as well as being the minimum-norm (i.e. minimum length) solute onto the PDES model among all the solutions in $\sigma = A^-b$ (Appendix B.2). It provides a mathematical approach to overcome the challenge in solving a PDES model with a non-full rank coefficient matrix.

Demonstration with the "MASS" Package in R

To demonstrate the M-P Approach in solving the PDES model, a linear equation system without full rank, we make use of the R library "MASS" [4]. This package includes a function named "*ginv*". It is devised specifically to calculate the Moore-Penrose generalized-inverse of a matrix. We used this function to calculate the Moore-Penrose generalized-inverse of the coefficient matrix A in the PDES smoking behavior model described in Equation (12).

As shown in Lin and Chen, smoking data from 2000 National Survey on Drug Use and Health (NSDUH) were compiled for US adsolescentsand young adults aged 15 to 21 (Table 1). According to the PDES, the state probability for each of the seven types of defined smokers by single year of age was calculated with the NSDUH data (Table 1). The state probabilities were estimated as the percentages of subjects in various behavioral states. Since the five smoking stages (i.e. NS, EX, SS, RS, QU) were all defined on the current year, the sum of them were one (i.e. 100%). While $\overline{SS}$ and $\overline{QU}$ were defined as the participants who self-stopped smoking and quit one year before.

With data for the first five types of smokers in Table 1, we estimated the transition probabilities with the M-P Approach. The results are presented in Table 2 (the R codes are included in Appendix A).

For validation and comparison purpose, we also computed the transitional probabilities using data for all seven types of smokers and the original PDES method by Lin and Chen using R (Codes are included also in Appendix A) [10]. The results from Table 3 were almost identical to those reported in the original study by Lin and Che

n. As we expected, by comparing the results in Table 2 with those in Table 3, for the five transitional probabilities (e.g., σ_1, σ_2, σ_7, σ_8, σ_9) that are not directly affected by the two additionally defined stages $\overline{SS}$ or old self-stoppers and $\overline{QU}$ or old quitters, the results from the "M-P Approach" are almost identical to those from the original method. On

Age	NS	EX	SS	RS	QU	$\overline{S}$	$\overline{QU}$
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 1: Percentages of People in 2000 NSDUH Smoking Data.

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	0.10	0.52	0.26	0.74	0.38	0.93	0.07	0.54	0.46
16	0.88	0.12	0.22	0.40	0.40	0.60	0.38	0.94	0.06	0.53	0.47
17	0.88	0.12	0.20	0.38	0.41	0.59	0.42	0.94	0.06	0.53	0.47
18	0.86	0.14	0.21	0.39	0.41	0.59	0.40	0.95	0.05	0.53	0.47
19	0.89	0.11	0.28	0.43	0.43	0.57	0.28	0.95	0.05	0.53	0.47
20	0.96	0.04	0.37	0.52	0.42	0.58	0.11	0.95	0.05	0.53	0.47

Table 2: Transitional probabilities of the PDES smoking model from "M-P Approach".

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	0.39	0.61
16	0.88	0.12	0.36	0.27	0.27	0.73	0.38	0.95	0.05	0.39	0.61
17	0.88	0.12	0.28	0.31	0.34	0.66	0.41	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	0.27	0.73
20	0.96	0.04	0.62	0.28	0.19	0.81	0.11	0.97	0.03	0.27	0.73

Table 3: Replication of the transitional probabilities of the PDES smoking model derived with the original method by Lin and Chen and data from the 2000 NSDUH but computed using R.

the contrary, however, the other six estimated probabilities (σ_3, σ_4, σ_5, σ_6, σ_{10}, σ_{11}) differed between the two methods. For example, compared with the original estimates by Lin and Chen, σ_{10} (the transitional probability to relapse to smoke again) with the "M-P Approach" are higher and σ_{11} (the transitional probability of remaining as quitters) arelower; furthermore, these two probabilities show little variations across ages compared to the originally reported results.

To the best of our understanding, the results from the "M-P Approach" are more valid for a number of reasons. (1) The M-P Approach did not use additional data from which more errors could be introduced. (2) More importantly, the results from the M-P Approachscientifically make more sense than those estimated with the original method. Using σ_{10} and σ_{11} as examples, biologically, it has been documented that it is much harder for adolescent smokers who quit and remain as quitters than to relapse and smoke again [13-15]. Consistent with this finding, the estimated σ_{10} (quitters relapse to regular smokers) was higher and σ_{11} (quitters remain as quitters) was lower with the new method than those with the original method. The results from the "M-P Approach" more accurately characterize these two steps of smoking behavior progression. Furthermore, the likelihood to relapse or to remain as quitter is largely determined by levels of addiction to nicotine, rather than chronological age [16-20]. Consistent with this evidence, the estimated σ_{10} and σ_{11} with the "M-P Approach" varied much less along with age than those estimated with the original method. Similar evidence, supporting a high validity of the "M-P Approach", is the difference in the estimated σ_6 (self-stoppers remaining as self-stoppers) between the two methods. The probability estimated through the "M-P Approach" showed a declining trend with age, reflecting the dominant influence of peers and society rather than nicotine dependence [13,21,22]. However, no clear age trend was observed in the same probability σ_6 estimated using the original method by Lin and Chen.

Evaluation of Intervention Impact for Smoking Behaviors

As indicated in the previous section, the introduction of the "M-P Approach" will greatly facilitate the application of the PDES method in behavior research. In addition to characterizing smoking behavior, and to assessing effects from exposure to prevention programs, the PDES method can be used to predict changes in smoking behavior in the future, supporting public health planning and decision-making [8,9,11]. Next, we introduce the "M-P Approach" and the PDES model to evaluate the intervention impact for smoking behaviors.

As seen from the PDES model, the multi-stage behavioral transitions provide information on the likelihood that a person will progress from never-smoking (NS) to start smoking (EX), further to regular smoking (RS); regular smokers can quit smoking (QU) and quitters may relapse and become regular smokers again. These transitional probabilities are influenced by the environment of 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 campaign, are intended to change the environment and hence the transitional probabilities. Different tobacco control programs have different impacts on the transitional probabilities. For example, restriction of tobacco sales to minors and school-based programs has 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*, this PDES can be employed for this purpose. We illustrate this evaluation of intervention impact both theoretically and numerically as follows.

Suppose new intervention program is devised to reduce $\sigma_i(a)$ to $\sigma'_i(a)$. Corresponding to equations (2) to (6), the new transition matrix with the multi-stage vector (NS, EX, RS, SS, QU) can be denoted by: .

$$\Pi'(a)=\begin{bmatrix} 1-\sigma'_2(a) & 0 & 0 & 0 & 0 \\ \sigma'_2(a) & 1-\sigma'_4(a)-\sigma'_7(a) & 0 & \sigma'_5(a) & 0 \\ 0 & \sigma'_7(a) & \sigma'_8(a) & 0 & \sigma'_{10}(a) \\ 0 & \sigma'_4(a) & 0 & 1-\sigma'_5(a) & 0 \\ 0 & 0 & \sigma'_9(a) & 0 & 1-\sigma'_{10}(a) \end{bmatrix} \quad (13)$$

Let the multi-stage transitional 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. Therefore, the future smoking behavior distribution at different ages can be calculated as follows:

$$\begin{cases} Q'(a+1)=\Pi'(a)Q'(a) \\ Q'(a+2)=\Pi'(a+1)\Pi'(a)Q'(a) \\ \dots \\ Q'(a+k)=\Pi'(a+k-1)\dots\Pi'(a+1)\Pi'(a)Q'(a) \end{cases} \quad (14)$$

Let's use the original 2000 NSDUH Smoking Data in Table 1 and the estimated transitional probabilities from "M-P Approach" in Table 2 to illustrate the program impact. Suppose a tobacco intervention program is designed to decrease the probability of σ_2 (i.e. "Never-Smoking (NS)" to "Experimenter (EX)") by 20%. This 20% reduction would change the estimated probabilities in Table 2 for σ_2 from (16.6%, 11.6%, 12.3%, 13.8%, 10.7%, 4.4%) to (13.3%, 9.3%, 9.8%, 11.0%, 8.6%, 3.5%) for age of 15, 16, 17, 18, 19 and 20, respectively. With this 20% reduction from the intervention program, the smoking multi-behavioral distribution can be calculated using equation (14) as seen in Table 4 as follows:

Table 4 can be compared to Table 1 to investigate the percent changes of the smoking population for each smoking behavior under different age. For example, we can investigate the absolute change of the smoking population using the differences between the values from Table 4 to Table 1 as well as the relative change (Table 5) using these differences rescaled to the values in Table 1. For example, with this

Age	NS	EX	SS	RS	QU
15	63.65	12.81	14.74	7.84	0.66
16	55.21	11.70	18.85	12.55	1.39
17	50.10	12.65	22.81	12.33	1.82
18	45.17	12.54	27.68	12.12	2.20
19	40.19	12.64	32.40	11.97	2.49
20	36.75	12.13	35.80	12.38	2.63
21	35.45	10.96	36.64	13.53	3.12

Table 4: Smoking behavior if 20% reduction of $\sigma_2(a)$ from an intervention tobacco control program.

Age	NS	EX	SS	RS	QU
16	4.0	-24.8	6.5	0.8	57.8
17	6.7	-23.6	34.2	-31.5	53.9
18	9.6	-22.2	68.8	-50.5	34.0
19	13.1	-20.4	103.9	-60.8	19.8
20	15.8	-19.6	123.1	-64.3	11.6
21	16.8	-20.0	113.0	-62.2	6.3

Table 5: Relative change (in %) of multi-behavioral smoking population with 20% reduction of $\sigma_2(a)$ from an intervention tobacco control program.

20% reduction, the smoking "experimenter (EX)" population would be changed from (15.57%, 16.56%, 16.11%, 15.89%, 15.09%, 13.69%) to (11.70%, 12.65%, 12.54%, 12.64%, 12.13%, 10.96%) for age 16, 17, 18, 19, 20 and 21, respectively. This is accountable for a (24.8%, 23.6%, 22.2%, 20.4%, 19.6%, 20.0%) relative reduction in the population of "Experimenter (EX)" as seen in Table 5 (in column "EX"). Furthermore, with this 20% reduction, the "never-smoker (NS)" population would increase by about 4% to 17%, the "self-stopper (SS)" by 6.5% for age 16, but will dramatically increase to 113% for age 21, the "quitter (QU)" from early age of 16 by 58% to 6.3% for age 21 as seen in Table 5. It is also interesting to see from Table 5 that the "regular smoker (RS)" dropped by 31% for age 17 by more than 60% for ages 19, 20 and 21.

Discussion and Conclusions

The Moore-Penrose generalized-inverse matrix theory has significant applications in many fields, including multivariate analysis, operations research, neural network analysis, pattern recognition, system control, and graphics processing [3-7]. To the best of our knowledge, this is the first time this "M-P Approach" is used in solving a PDES model to describe smoking behavior progression in an adolescent population. Our study fills a methodology gap in PDES modeling. After an introduction to the "M-P Approach", we illustrate its application with the same data reported in the original study using the R software [10]. Results from the analysis using the "M-P Approach", although using less data, better reflect the dynamics of smoking behavior change in adolescents than do the results from the original analysis.

Findings of this study provide evidence that the "M-P Approach" can be used to solve a PDES model constructed to characterize complex health behaviors with cross-sectional data even if the coefficient matrix has no full rank. Behavioral modeling, like in many other systems research fields, has frequently been challenged because of the lack of "fully" observed data to quantitatively characterize a system, even when the system is constructed based on scientific theory or data. Successful application of the "M-P Approach" in solving the PDES model for smoking behavior will greatly facilitate system modeling of various other human behaviors with or without fully observed data.

According to the "M-P Approach", as long as a model is "true" (e.g., as long as it has a solution), it should be solvable even with partial observations. In our study, since the PDES smoking model has been proved to be true through previous analysis, the "M-P Approach" works. This success is not by chance. Similar to a system with extra observed data (e.g., multiple regression with the number of equations greater than the number of unknowns) that can be solved using the "M-P Approach" (e.g., the least square approach is in theory a "M-PApproach"), a system with the number of unknowns greater than the number of independent equations (e.g., partially observed data) can also be solved based on the minimum-norm approach with M-P inverse matrix.

Despite a successful application of the M-P approach in solving a PDES model, more research is needed to investigate more specific conditions in which the application of the "M-P Approach" is indicated to solve complex modeling questions with a linearequation system but without a full-rank coefficient matrix. We are initializing a systematic simulation study to validate this new approach.

Acknowledgement

References

1. Moore EH, Barnard RW (1935) General analysis Parts I, Philadelphia: The American Philosophical Society.

2. Penrose RA (1955) generalized inverse for matrices Procedings of the Cambridge Philosophical Society 51: 406-413.

3. Campbell SL, Meyer CD (1979) Generalized inverses of linear transformations. London: Pitman

4. VenablesWN, Ripley BD (2002) Modern Applied Statistics with S, (4thedn) Springer New York.

5. Ying Z, Jia SH (2009) Moore-Penrose generalized inverse matrix and solution of linear equation group. Mathematics in practice and theory 39: 239-244.

6. Nashed MZ (1976) Generalized inverses and applications. New York, Academic press.

7. Cline RE (1979) Elements of the theory of generalized inverses for matrices. Modules and monographis in undergraduate mathematics and its application project.

8. Chen X, Lin F (2012) Estimating transitional probabilities with cross-sectional data to assess smoking behavior progression: A validation analysis. J Biometrics and Biostatistics.

9. Chen X, Lin F, Zhang X (2010) Validity of PDES Method in Extracting Longitudinal Information from Cross-Sectional Data: An Example of Adolescent Smoking Progression. American Journal of Epidemiology.

10. Lin F, Chen X (2010) Estimation of Transitional Probabilities of Discrete Event Systems from Cross-Sectional Survey and its Application in Tobacco Control. Inf Sci (Ny) 180: 432-440.

11. Chen X (2012) Exposure to school and community based prevention programs and reductions in cigarette smoking among adolescents in the United States, 2000-08. Evaluation and Program Planning 35: 321-328.

12. Penrose R (1955) A generalized inverse for matrices. Proceedings of the Cambridge Philosophical Society 51: 406-413.

13. Turner LR, Veldhuis CB, Mermelstein R (2005) Adolescent smoking: Are infrequent and occasional smokers ready to quit? Substance Use and Misuse 40: 1127-1137.

14. Kralikova E (2013) Czech adolescent smokers: unhappy to smoke but unable to quit. International Journal of Tuberculosis and Lung Disease 17: 842-846.

15. Reddy SR, Burns JJ, Perkins KC (2003) Tapering: An alternative for adolescent smokers unwilling or unable to quit.

16. Panlilio LV (2012) Novel use of a lipid-lowering fibrate medication to prevent nicotine reward and relapse: preclinical findings. Neuropsychopharmacology 37: 1838-1847.

17. Carmody TP (1992) Preventing relapse in the treatment of nicotine addiction: current issues and future directions. J Psychoactive Drugs 24: 131-158.

18. Drenan RM, Lester HA (2012) Insights into the neurobiology of the nicotinic cholinergic system and nicotine addiction from mice expressing nicotinic receptors harboring gain-of-function mutations. Pharmacol Rev 64: 869-879.

19. Govind AP, Vezina P, Green WN (2009) Nicotine-induced upregulation of nicotinic receptors: underlying mechanisms and relevance to nicotine addiction. Biochem Pharmacol 78: 756-765.

20. DeBiasi M, Salas R (2008) Influence of neuronal nicotinic receptors over nicotine addiction and withdrawal. Exp Biol Med (Maywood): 917-929.

21. Lim MK (2012) Role of quit supporters and other factors associated with smoking abstinence in adolescent smokers: a prospective study on Quitline users in the Republic of Korea. Addict Behav 37: 342-345.

22. Castrucci BC, Gerlach KK (2005) The association between adolescent smokers' desire and intentions to quit smoking and their views of parents' attitudes and opinions about smoking. Matern Child Health J 9: 377-384.