IIE Transactions (2015) **47**, 1–10

ISSN: 0740-817X print / 1545-8830 online
DOI: 10.1080/0740817X.2015.1009197

Modeling dependence in health behaviors

BRANDON POPE[1], ABHIJIT DESHMUKH[2], ANDREW JOHNSON[3] and J. JAMES ROHACK[4]

[1]*Baylor Scott & White Health, Dallas, TX 75246, USA*
E-mail: brandonpope84@gmail.com
[2]*School of Industrial engineering, Purdue University, West Lafayette, IN 47907, USA*
[3]*Department of Industrial and Systems Engineering, Texas A&M University, College Station, TX 77843, USA*
[4]*Baylor Scott & White Center for Healthcare Policy, Temple, TX 76508, USA*

Received December 2011 and accepted December 2014

The prediction and control of distributed healthcare behaviors within a population such as smoking, diet, and physical activity are of great concern to those who pay for healthcare, including employers, insurers, and public policy makers, given the significant effect on costs. In considering the selection of multiple health behaviors, the nature of dependence between behaviors must be considered because simplifying assumptions such as independence are untenable. Using data from the National Heart, Lung, and Blood Institute, we find strong evidence to reject the hypothesis of independence between the aforementioned behaviors, while finding some evidence of conditional independence. In this article, several alternatives to the assumption of independence are presented, each of which significantly improves the ability to predict combined behavior. We present models of dependence through marginal probabilities and, taking inspiration from non-expected utility maximizing behavior, through attractions to behavioral alternatives. We find that consistently healthy (or unhealthy) combinations of behaviors are more likely to occur relative to the assumption of independence. We discuss how our results could be used in designing policies to curtail costs and improve health.

Keywords: Conditional dependence, copulas, health behaviors

1. Introduction

The current unsustainable trend in healthcare expenditures has produced an explosion of research into the processes, operations, and techniques of healthcare service providers. However, relatively less attention has been given to the distributed decisions of healthcare consumers that have implications for costs as well as quality of life. For example, the cost of physical inactivity alone was estimated to be $76.6 billion (USD) in the year 2000 (Pratt *et al.*, 2000). The prediction and control of healthcare behaviors such as smoking, diet, and physical inactivity are of interest to those who pay for healthcare, including employers, insurers, and policy makers, given their large effect on costs. These three behaviors are of particular interest given their influence in the prevention and management of chronic diseases such as coronary heart disease. These tasks are particularly challenging due in part to the difficulty in developing accurate models of consumer behavior in healthcare situations. A common normative assumption made to characterize consumer behavior is expected utility maximization. However, the information and intelligence requirements imposed on healthcare consumers by the assumption of expected utility maximization make it an implausible description of behavior. Identifying and forecasting health outcomes, estimating the costs of those outcomes, and weighing the impacts of personal health decisions are highly complex problems facing healthcare consumers.

From a single consumer's internal perspective, health behavior is a decision and hence not uncertain. However, for a policy maker or payer considering a population of consumers, the health behavior of a consumer can be modeled as a random variable. When considering multiple behaviors, care must be taken to consider the nature of the dependence between the behaviors. The boundaries of dependence, total dependence, and complete independence are the most frequently used simplifying assumptions to model these relationships but are often used primarily for their analytical or computational convenience, rather than their realism or accuracy (Ferson *et al.*, 2004). In many circumstances these assumptions are doubtful, unlikely, or clearly wrong. Health-related behaviors of healthcare consumers is one such example where these simplifying assumptions are untenable.

*Corresponding author

In this article, several alternatives to the simplifying assumption of independence between behaviors are presented to model the healthcare consumers diet, exercise, and smoking behaviors. The first models presented model the dependence through conditional probabilities and our empirical finding of conditional independence between smoking and dietary behaviors. The second approach to modeling dependence is inspired by the lack of expected utility maximization on the behalf of healthcare consumers. A common approach to modeling decision making in such scenarios is through the use of learning rules (Fudenberg and Levine, 1998), which model attractions to the various decision alternatives. When the marginal distributions over behaviors are governed by attractions, as in the case of learned behaviors, dependence can be modeled through computed probabilities or through the attractions themselves. To model dependence through attractions to multiple behaviors, we introduce the concept of a joint attraction function. To illustrate the importance of considering dependence, and to compare the proposed models of dependence, we use data from the National Heart, Lung, and Blood Institute (NHLBI). The results show that independence between behaviors clearly fails and that the alternatives presented significantly improve prediction of the combined diet, exercise, and smoking behavior.

The remainder of this article is organized as follows: Section 2 reviews the literature on modeling probabilistic dependence and the use of learning rules to model behavior. Section 3 introduces the data and examines the assumption of independence between behaviors. Section 4 presents several alternatives for modeling dependence between behaviors, first by conditional probabilities and, second, by joint attraction functions. Cross-validation of the models shows that each significantly reduces prediction error relative to the assumption of independence. Section 5 presents a longitudinal validation of the approaches, and Section 6 discusses our findings, their relevance for healthcare policy, and directions for future research.

2. Literature review

Dependence between random variables is an old yet important concept for any discipline considering multiple random variables. See Lehmann (1966) for an early discussion of dependence concepts, measures of dependence, and tests for independence. More recently, Ferson *et al.* (2004) provided an excellent review of dependence modeling, including a detailed look at the relationship between dependence and correlation. Application domains modeling and illustrating dependencies include medicine (Chessa *et al.*, 1999), reliability (Vesely *et al.*, 1981), geological exploration (Keefer, 2004; Bickel and Smith, 2006), and risk assessment (Clemen and Reilly, 1999). Copulas (Nelsen, 2006) have been the most popular means of specifying a dependence structure given two or more continuous marginal distributions. However, we avoid their use here since we are interested in discrete decisions, amenable to modeling via learning rules, with discontinuous marginal distributions. Although the behaviors we study could be defined and measured on continuous scales, it would be unclear which copula family to assume. We prefer simpler descriptions of dependence that seem more plausible given the behavioral health preferences of consumer, which drive dependencies in health behaviors. Decomposing a joint probability distribution using Bayes' Rule to construct dependence trees is a well-established approach (Chow and Liu, 1968) for approximation and estimation of a given data set. The value in constructing a model of this conditional dependence is the ability to apply the model outside the derivation data set and to better understand the underlying process of dependence. Both objectives would be desirable in the health behavior domain.

Our second approach to modeling dependence is inspired by non-expected utility models of preferences. The departure from the normative assumption of expected utility maximization is motivated by the arguments of bounded rationality (Simon, 1955; Ellison, 2006), which have given rise to various non-expected utility models of decision making (Machina, 2004). Among these, learning rules are intuitive and cognitively plausible and describe how choices and preferences dynamically evolve in individual decision frameworks and games (Fudenberg and Levine, 1998). Flexible learning models such as the experience-weighted attraction model (Camerer and Ho, 1999) generalize various learning protocols such as choice reinforcement models (Roth and Erev, 1995) and belief-based models (e.g., fictitious play (Brown, 1951)). The basic setup is that facing a repeated decision from a finite set of alternatives $\mathcal{A}$, an individual chooses an alternative based on attractions ($\{s_i\}_{i\in\mathcal{A}} = s \in \mathcal{S}$) to each alternative and receives a payoff or reward $x \in \mathcal{X}$ at each decision epoch. A learning rule, $\mathcal{L} : \mathcal{S} \times \mathcal{A} \times \mathcal{X} \to \mathcal{S}$, computes the subsequent epoch's attraction given the previous attractions, alternative chosen, and payoffs received. We note, however, that the dynamic learning rules of the economic community have yet to be incorporated into the behavioral health domain. Psychology-based models such as the Theory of Planned Behavior (TPB); Ajzen (1991) analyze the factors that comprise behavioral intent and have been applied with success in the health domain (Godin and Kok, 1996). However, TPB does not explain how these factors evolve over time. There is a scarcity of research regarding how multiple behaviors are dependently chosen from the learning perspective. The existing work on multiple health behavior modification has focused on intervention design (Glasgow *et al.*, 2004). We now turn to the data and what insights can be gleaned regarding dependence between behaviors.

3. Data examination

Formally, the three behaviors considered in this article are binary decisions about diet, exercise, and smoking. For each behavior, the healthy choice will be denoted by a one,

Table 1. Joint behaviors in ARIC data, ($pr(E_1) = 0.46$, $pr(D_1) = 0.24$, $pr(S_1) = 0.79$)

	E_0			E_1	
	S_0	S_1		S_0	S_1
D_0	1054	3323	D_0	641	2817
D_1	264	913	D_1	207	1090

whereas the unhealthy choice will be denoted by zero. These zero/one distinctions are based on the guidelines of the American Heart Association (AHA), to not smoke (S_1; smoking $= S_0$), get at least 150 minutes of moderate physical activity per week (E_1; $< 150 = E_0$), and to consume a diet with less than 10% of total calories coming from saturated fat (D_1; $\geq 10\% = D_0$).

The data set used in this article is from the Atherosclerosis Risk In Communities (ARIC) Study conducted by the NHLBI. The study followed cohorts of individuals from four geographically diverse communities, tracking various aspects of cardiovascular health and behavior. Although the longitudinal aspect is not crucial for this article, the data set does provide detailed information to measure behavior relative to the AHA guidelines. Using data from a single time period (so that independence between the observations is reasonably assumed) provides behavioral observations of 10 309 individuals. Table 1 provides a summary table of the data.

3.1. *Independence*

To test for independence between pairs of behaviors, χ^2 tests were performed. For $n \times m$ contingency tables, the test statistic $\chi^2 = \sum_{i,j}(O_{ij} - \widehat{\mathbb{E}}_{ij})^2/\widehat{\mathbb{E}}_{ij}$, where O_{ij} is the observed frequency of cell ij, and $\widehat{\mathbb{E}}_{ij}$ is the expected frequency of cell ij based on the marginal probabilities of i and j and has the χ^2 distribution with $(n-1)(m-1)$ degrees of freedom (Agresti, 2007). For each pair of behaviors, three tests of independence were performed. The test was performed within subsets of the population defined by choice in the third behavior (0/1), and combined (C) across the third behavior. The results displayed in Table 2 show that the relationship between diet and smoking is the only pair for which the null hypothesis of independence is not easily rejected. These results from the χ^2 test were compared and found to be quite similar to other tests for independence such as the G-test (see results in the Appendix).

To further clarify the dependence structure observed in the data, the Cochran–Mantel–Haenszel (CMH) test (Gastwirth, 1984) was performed to check for conditional independence. In data where there are more than two variables, this test checks for conditional independence based on the strength of the association measured by the odds ratio. The test results are summarized in Table 3. The CMH test fails to reject the hypothesis that conditioned on the level of exercise, diet and smoking behavior are statistically independent, since the null hypothesis fails to be rejected at the 0.05 significance level with the Bonferroni correction. The odds ratio pooled over levels of the conditioning variable also indicates a relatively strong association between exercise and both smoking and dietary behavior. We use these results to inform our modeling of dependence, described in the next section.

Table 2. Pairwise χ^2 tests for independence between diet (D), smoking (S), and exercise (E)

Test pair	*Third behavior level*	χ^2	*p-Value*
DS	0	1.30	0.253
DS	1	4.10	0.043
DS	C	7.48	0.006
ES	0	34.69	0.000
ES	1	16.34	0.000
ES	C	53.32	0.000
ED	0	5.56	0.018
ED	1	43.78	0.000
ED	C	51.67	0.000

4. Modeling the dependence

The purpose of modeling the dependence between behaviors is to predict the probabilities for each type of combined behavior from marginal probabilities (equivalently separate attractions) to each behavior. We consider two ways to incorporate dependence: through the probabilities using the insights concerning conditional independence discovered in Section 3 or through the attractions via joint attraction functions. Figure 1 illustrates the dependence pathways.

4.1. *Models of conditional independence*

The first pathway we explore to model the dependence is directly through the probabilities, labeled ① in Fig. 1. The χ^2 tests from Section 3 showed that independence does not hold and hence marginal probabilities cannot simply be multiplied together to compute a joint probability for a combined behavior. However, the CMH tests provided evidence that conditional on exercise, diet and smoking are independent. Using this evidence allows simplified computation of the joint probability through Bayes' Rule:

$$\begin{aligned} pr(E_i D_j S_k) &= pr(E_i) \times pr(D_j S_k | E_i) \\ &= pr(E_i) \times pr(D_j | E_i) \times pr(S_k | E_i). \end{aligned} \quad (1)$$

Table 3. Results of the CMH tests for conditional pairwise independence between diet (D), smoking (S), and exercise (E)

Test pair	*CMH statistic*	*p-Value*	*Pooled OR*
DS	5.126	0.024	1.174
ES	51.133	0.000	1.434
ED	49.417	0.000	1.395

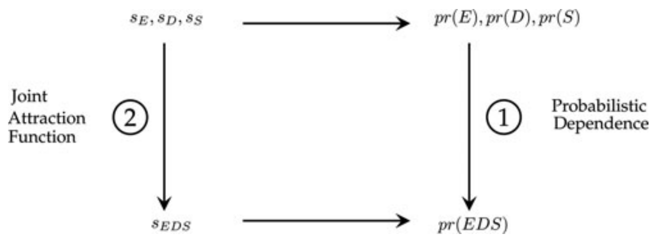

Fig. 1. Dependence pathways from behaviors attractions to the joint probability.

The question then becomes how to model $pr(D_j|E_i) \neq pr(D_j)$ or $pr(S_k|E_i) \neq pr(S_k)$. Our first approach is an additive effects model.

4.1.1. *Conditional independence–additive dependence*

The intuition behind this model is that exercise decisions would tend to support or reinforce similar diet and smoking decisions by increasing the probability of those similar decisions. Here we model this increase through additive functional forms and parameters, α:

$$\begin{array}{ll} pr(D_1|E_1) = pr(D_1) + \alpha_{D_1|E_1} & pr(S_1|E_1) = pr(S_1) + \alpha_{S_1|E_1} \\ pr(D_0|E_1) = pr(D_0) - \alpha_{D_1|E_1} & pr(S_0|E_1) = pr(S_0) - \alpha_{S_1|E_1} \\ pr(D_1|E_0) = pr(D_1) + \alpha_{D_1|E_0} & pr(S_1|E_0) = pr(S_1) + \alpha_{S_1|E_0} \\ pr(D_0|E_0) = pr(D_0) - \alpha_{D_1|E_0} & pr(S_0|E_0) = pr(S_0) - \alpha_{S_1|E_0} \end{array}.$$

All of the α values here are easily estimated from the data by differencing observed probabilities (e.g., $\widehat{\alpha}_{D_1|E_1} = \widehat{pr}(D_1|E_1) - \widehat{pr}(D_1)$). Performing these computations on the ARIC data displayed in Table 1 leads to $\widehat{\alpha}_{D_1|E_1} = 0.0328$, $\widehat{\alpha}_{S_1|E_1} = 0.0318$, $\widehat{\alpha}_{D_1|E_0} = -0.0281$, and $\widehat{\alpha}_{S_1|E_0} = -0.0272$. Note that $\widehat{\alpha}_{D_1|E_1} \approx \widehat{\alpha}_{S_1|E_1} \approx -\widehat{\alpha}_{D_1|E_0} \approx -\widehat{\alpha}_{S_1|E_0} \approx 0.03$. With these parameter estimates, the resulting conditional probabilities will only be meaningful if the marginal probabilities of diet or smoking fall in the range [0.03, 0.97]. Based on these α values, the Conditional Independence–Additive Dependence (CIAD) model can be used to predict occurrences of each type of behavior. Evidence for the goodness-of-fit of this model is presented after introducing our second model of dependence.

4.1.2. *Conditional independence–multiplicative dependence*

Another approach for modeling dependence based on the same intuition is a multiplicative approach, $pr(D|E) = \beta_{D|E} \times pr(D)$. The β parameters represent multiplicative adjustments to likelihoods based on the relationship between diet (or smoking) and exercise behaviors. This approach could be viewed as an analog to Keefer's underlying event model (Keefer, 2004), where exercise is the "underlying behavior" driving the combined behavior. To ensure meaningfulness of the probabilities, note that both

Table 4. Comparing conditional independence models for predicting joint behaviors

Behavior	*Data*	*IND*	*CIAD*	*CIMD*
000	1054	888	1055	1026
001	3323	3342	3342	3365
010	264	281	281	275
011	913	1055	888	901
100	641	757	623	618
101	2817	2847	2839	2842
110	207	239	230	229
111	1090	899	1050	1053
SSE	—	100 143	4199	5798

the data and the additive approach indicate that when $i \neq j$, $pr(D_i|E_j) < pr(D_i)$ and the analogous relationship between exercise and smoking. This results in an estimated value of $\beta < 1$ when $i \neq j$. This insight proves useful since a probability multiplied by a scalar less than one is still a probability. Using this insight, we use four parameters, β, to model decreases in conditional probabilities:

$$\begin{array}{ll} pr(D_1|E_1) = 1 - pr(D_0|E_1) & pr(S_1|E_1) = 1 - pr(S_0|E_1) \\ pr(D_0|E_1) = \beta_{D_0|E_1} \times pr(D_0) & pr(S_0|E_1) = \beta_{S_0|E_1} \times pr(S_0) \\ pr(D_1|E_0) = \beta_{D_1|E_0} \times pr(D_1) & pr(S_1|E_0) = \beta_{S_1|E_0} \times pr(S_1) \\ pr(D_0|E_0) = 1 - pr(D_1|E_0) & pr(S_0|E_0) = 1 - pr(S_1|E_0) \end{array}.$$

Estimating the β values from the ARIC data in Table 1 results in $\widehat{\beta}_{D_0|E_1} = 0.96$, $\widehat{\beta}_{D_1|E_0} = 0.88$, $\widehat{\beta}_{S_0|E_1} = 0.85$, and $\widehat{\beta}_{S_1|E_0} = 0.97$. Our ability to identify dissimilar behaviors ensured this Conditional Independence–Multiplicative Dependence (CIMD) model would produce meaningful probabilities, although in general, care would need to be taken to ensure meaningful probabilities.

Table 4 reports the sum of squared differences between observed outcomes and predicted outcomes for both the CIAD and CIMD models. These are compared with the alternative model of assuming independence (IND). The left-most column indicates the combined behavior, represented by the exercise behavior in the hundreds digit, diet behavior in the tens digit, and smoking behavior in the ones digit. The second column from the left reports the number of observations found in the data for each behavior, as shown in Table 1. The overall Sum of Squared Error (SSE) is reported in the bottom row; smaller values indicate better predictive power.

4.1.3. *Validation of conditional independence models*

In order to validate the models of conditional independence, we divide our data set into training (75%) and testing (25%) subsamples. Table 5 reports the mean and standard error of the SSEs in the testing data sample with parameters estimated from the training sample over 1000 subsamples. In order to provide a comparison with extant methods in

Table 5. Standard errors on testing data by dependence models and copulas

	IND	*CIAD*	*CIMD*	C_F	C_C
$\overline{SSE}$	6959.8	1207.6	1150.2	1721.9	1741.8
SE_{SSE}	68.1	28.8	26.1	35.9	35.4

the literature, we also model dependence through copulas. From Nelsen (2006) we estimate the dependence by two comprehensive, Archimedian copulas, Frank's copula:

$$C_F(u, v) = -\frac{1}{\theta} \ln \left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{(e^{-\theta} - 1)} \right),$$

and Clayton's copula:

$$C_C(u, v) = \left[\max\{u^{-\theta} + v^{-\theta} - 1, 0\} \right]^{-1/\theta}.$$

The dependence parameter, θ, was estimated for each copula by the maximum likelihood method.

Each conditional independence model predicts joint behaviors significantly better than pure independence. Our results show that both CIAD and CIMD models reduce the SSEs by at least 80% and are statistically indistinguishable from each other. Both copulas also significantly improve over the independence assumption, whereas neither copula reduces errors as much as the CIMD or CIAD models. In terms of pros and cons that could aid in distinguishing between CIAD and CIMD, there are few. Both models function similarly based on the intuition that the likelihood of a dependent behavior will increase (decrease) when a similar (dissimilar) behavior is chosen. When the populations for which the models are estimated and applied have similar marginal probabilities (as would typically be the case), the models will behave equivalently.

Although our models of dependence were presented in the context of binary decisions, each could be generalized to the case of three or more alternatives per decision, albeit at the increased cost of parameters needing to be estimated. One approach to dealing with the increasing number of parameters in the case of greater alternatives would be to impose additional assumptions regarding the form of dependence. For example, consider modeling diet and exercise behavior with N alternatives each, again ranging from unhealthy to healthy. An additive dependence model where diet behaviors are more likely when conditioned on exercise behaviors of similar healthiness could make the assumption that all other behaviors are equally less likely:

$$pr(D_i|E_j) = \begin{cases} pr(D_i) + \alpha_j & i = j \\ pr(D_i) - \alpha_j/(N-1) & i \neq j \end{cases}.$$

Another reasonable assumption is that increasingly different behaviors will be increasingly less likely to be observed jointly, which could be modeled by a multiplicative dependence function,

$$pr(D_i|E_j) = \begin{cases} \beta_{D|E}^{|i-j|} pr(D_i) & i \neq j \\ 1 - \sum_{k \neq i} pr(D_k|E_j) & i = j \end{cases}.$$

Having presented two models of dependence through pathway ①, next we describe modeling dependence via pathway ② using joint attraction functions.

4.2. *Joint attraction functions*

We now take the approach of modeling consumer preferences through attractions, s_a, to various alternatives, $a \in \mathcal{A}$. These attractions can be interpreted as expected utilities in the sense that alternatives with higher attractions are preferred. As mentioned earlier, the attractions approach would be desirable when modeling the evolution of preferences through a learning rule. The relationship between attractions and probabilities is computed through a Stochastic Choice Function (SCF), $f : \mathcal{S} \to \Delta(\mathcal{A})$, which maps behavior alternative attractions to the unit simplex on alternatives. We limit our focus to the logit rule, where given real-valued attractions $\{s_a\}_{a \in \mathcal{A}}$, the probability of choosing an alternative, a, is computed by

$$pr(a) = \frac{e^{\lambda \cdot s_a}}{\sum_{a' \in \mathcal{A}} e^{\lambda \cdot s_{a'}}}.$$

The logit rule is non-unique up to a normalization of the attractions, which allows preferences to binary behaviors to be represented using a single scalar value (with the other attraction normalized to zero). We will now assume that individuals form attractions to joint behaviors by combining attractions to separate behaviors using a *joint attraction function*. This assumption is relatively benign, implying only that a consumer's preferences over a joint decision are a function of the consumer's preferences over each individual decision. A joint attraction function takes as inputs attractions $(\{s^i_{a(i)}\}^n_{i=1})$ to alternatives of multiple behaviors $(i = 1, \ldots, n)$ and computes a joint attraction to the combined behavior, $a = (a(1), a(2), \ldots, a(n))$. For example, the joint attraction function could be simple addition, multiplication, or a combination. The functional form of the joint attraction function through the stochastic choice rule defines a relationship of dependence between behaviors.

Proposition 1. *Under the logit rule, a simple additive joint attraction function*

$$g\left(s^1_{a(1)}, s^2_{a(2)}, \ldots, s^n_{a(n)}\right) = \sum_{i=1}^{n} s^i_{a(i)}$$

is equivalent to assuming independence between behavioral probabilities.

Table 6. Additive and interaction joint attraction functions for predicting joint behaviors

Behavior	Data						
	θ_E =	*1*	*0.97****	*1.47****	*2.75****	*3.25****	*3.20**
	θ_D =	*1*	*1.00****	*1.14****	*1.00****	*1.14****	*1.23***
	θ_S =	*1*	*1.00****	*1.00****	*0.88****	*0.88****	*0.80***
	θ_{ED} =	*0*	*0*	*5.31****	*0*	*5.31***	*5.19**
	θ_{ES} =	*0*	*0*	*0*	*−5.00****	*−5.00***	*−4.90**
	θ_{DS} =	*0*	*0*	*0*	*0*	*0*	*−0.26*
000	1054	888	886	919	1002	1039	1060
001	3323	3342	3335	3457	3220	3338	3317
010	264	281	280	247	316	279	258
011	913	1055	1053	930	1017	898	919
100	641	757	759	726	644	617	635
101	2817	2847	2855	2731	2969	2841	2823
110	207	239	240	272	204	231	213
111	1090	899	902	1025	938	1066	1084
SSE	—	100 143	100 001	59 925	73 328	3296	329

Proof.

$$
\begin{aligned}
&pr[a(1), a(2), \ldots, a(n)] \\
&= \frac{e^{\lambda\cdot g(s^1_{a(1)} s^2_{a(2)} \cdots s^n_{a(n)})}}{\sum_{j(1)\in\mathcal{A}^1}\cdots\sum_{j(k)\in\mathcal{A}^n} e^{\lambda\cdot g(s^1_{j(1)} s^2_{j(2)}\cdots s^n_{j(n)})}} \\
&= \frac{e^{\lambda\cdot\sum_{i=1}^n s^i_{a(i)}}}{\sum_{j(1)\in\mathcal{A}^1}\cdots\sum_{j(n)\in\mathcal{A}^n} e^{\lambda\cdot\sum_{i=1}^n s^i_{j(i)}}} \\
&= \frac{\prod_{i=1}^n e^{\lambda\cdot s^i_{a(i)}}}{\sum_{j(1)\in\mathcal{A}^1}\cdots\sum_{j(n)\in\mathcal{A}^n}\prod_{i=1}^n e^{\lambda\cdot s^i_{j(i)}}} \\
&= \frac{\prod_{i=1}^n e^{\lambda\cdot s^i_{a(i)}}}{\prod_{i=1}^n\left(\sum_{j(i)\in\mathcal{A}^i} e^{\lambda\cdot s^i_{j(i)}}\right)} \\
&= \prod_{i=1}^n \frac{e^{\lambda\cdot s^i_{a(i)}}}{\left(\sum_{j(i)\in\mathcal{A}^i} e^{\lambda\cdot s^i_{j(i)}}\right)} \\
&= \prod_{i=1}^n pr[a(i)].
\end{aligned}
$$

This completes the proof. ■

4.2.1. *Additive relaxation*

Proposition 1 will be used in conjunction with our findings from Section 3 to inform what types of joint attraction functions should be considered. Proposition 1 showed that independence is equivalent to a purely additive relationship from s_i^E, s_j^D, s_k^S to s_{ijk}^{EDS} in the context of the logit rule. Hence, relaxing this assumption by adding coefficients and cross-terms to the pure additive form will relax the assumption of independence. We use the general model

$$
\begin{aligned}
s_{ijk}^{EDS} = {} & \theta_E s_i^E + \theta_D s_j^D + \theta_S s_k^S + \theta_{ED} s_i^E s_j^D + \theta_{ES} s_i^E s_k^S \\
& + \theta_{DS} s_j^D s_k^S + \theta_{EDS} s_i^E s_j^D s_k^S, \qquad (2)
\end{aligned}
$$

to estimate the parameters $\theta = (\theta_E, \theta_D, \theta_S, \theta_{ED}, \theta_{ES}, \theta_{DS})'$, by an Ordinary Least Squares (OLS) approach. The results are displayed in Table 6, where p-values are computed using the Wald test from the observed Fisher information, and significance is denoted by *** for the 0.01 level, ** for the 0.05 level, and * for the 0.10 level. The right-most column of Table 6 shows the most flexible model with coefficients for all two-way interactions. However, the interaction between diet and smoking is insignificant, consistent with our previous findings. Therefore, this interaction term was dropped from the final joint attraction function model, shown in column 7 of Table 6. Columns 3 to 6 illustrate how progressively relaxing the pure additive assumption of independence increasing the predictive ability of the joint attraction function.

Although this linear joint attraction function with interactions (2) (herein JAF1) appears to capture dependencies to predict the observed behavior, the parameters estimated are difficult to interpret. In other words, what is the interpretation of a positive coefficient for the interaction between exercise and diet but a negative coefficient between exercise and smoking? In order to obtain more interpretable results, we introduce another joint attraction function.

4.2.2. *Behavioral consistency*

Results from our models of conditional independence showed the increased likelihood of consistently healthy (unhealthy) behaviors relative to independence. Building on these previous results, and seeking to improve on the interpretation of the model parameters, we also estimate the following joint attraction model:

$$
s_{ijk}^{EDS} = \theta_E s_i^E + \theta_D s_j^D + \theta_S s_k^S + \theta_{ED}\mathbb{1}_{i=j} + \theta_{ES}\mathbb{1}_{i=k}. \qquad (3)
$$

This model is designed to capture the increased attraction of (un)healthy smoking and diet behaviors when exercise behavior is (un)healthy. Equation (3) achieves this by increasing the joint attraction to behavior ijk by θ_{ED}

Table 7. Behavioral consistency joint attraction functions for predicting joint behaviors

Behavior	*Data*	θ_E 1, θ_D 1, θ_S = 1, θ_{ED} 0, θ_{ES} 0	1.09**, 1.00***, 1.02***, 0.06**, 0.06**
000	1054	888	1039
001	3323	3342	3339
010	264	281	279
011	913	1055	898
100	641	757	617
101	2817	2847	2841
110	207	239	231
111	1090	899	1066
SSD	–	100 143	3292

when exercise and diet behaviors are consistent (i.e., when $i = j$) and increasing the joint attraction by θ_{ES} when exercise and smoking behaviors are consistent (i.e., when $i = k$). We again estimate parameters by the approach OLS and the results are reported in Table 7, where now $\theta = (\theta_E, \theta_D, \theta_S, \theta_{ED}, \theta_{ES})'$.

These results show that joint behaviors that are consistently healthy (or unhealthy) receive attraction boosts through the individual's dependence mechanism relative to their inconsistent alternatives. This second joint attraction function (3) (herein JAF2) matches the predictive power of JAF1 (2) while producing more easily interpreted results. For both of the joint attraction functions presented we have found the predictive ability to be robust to the attraction normalization and SCF parameter, λ, chosen. The possibility remains that other nonlinear joint attraction functions could be more sensitive to the normalization and SCF chosen.

4.2.3. *Validation of joint attraction functions*

Similar to before, we validate the joint attraction function models by splitting our data set into training (75%) and testing (25%) subsamples. Table 8 reports the mean and standard error of the SSEs of squared errors in the testing data sample with parameters estimated from the training sample over 1000 subsamples. Each joint attraction function reduces the error of the independence-equivalent joint attraction function by over 50%. The second functional form (3), based on behavioral consistency, significantly outperforms the first, with a mean reduction in error of 69% relative to independence. Again the performance of Frank's and Clayton's copula models of dependence are shown for comparison. Both copulas appear to fit the dependence patterns better than the joint attraction approach.

Table 8. Standard errors on testing data by joint attraction functions and copulas

	IND	*JAF1*	*JAF2*	C_F	C_C
$\overline{SSE}$	6895.5	3056.8	2118.3	1721.9	1741.8
SE_{SSE}	66.8	63.3	32.3	35.9	35.4

4.3. *Longitudinal validation of dependence models*

We used each model {CIAD, CIMD, JAF1, JAF2} with parameters estimated from the ARIC data in Table 1 to predict joint behaviors from a second time period in the ARIC data. Our results appear in Table 9. The only model that does not improve on the assumption of independence is JAF1. This places some doubt as to the generality of the functional form and estimates of this approach. The other three models predict joint behaviors with significantly less error than independence. The second joint attraction function based on behavioral consistency (3) and the multiplicative conditional probability model each have less than a third of the error of independence, whereas the additive conditional probability model has less than half of the error. Although our cross-validation exercise from Table 5 could not distinguish between the two conditional independence models, these results indicate that the CIMD model may be a more generalizable model of dependence than CIAD. The copula perform the best of all of the tested models; one explanation for this is that these single-parameter approaches are less prone to over-fitting.

5. Simulation study

To study the properties of the dependence models over a broader range of possible binary outcomes and correlations, we conducted a simulation study. In order to simulate multi-dimensional binary outcomes, we used the method of Emrich and Piedmonte (1991), which has the ability to generate a wide range of positive and negatively correlated data (Chaganty and Joe, 2006). For simplicity and comparability with the results generated from the ARIC data set, we simulated data sets of three correlated binary outcomes with 10 000 observations. A randomly selected mean vector $(p_1, p_2, p_3) \in [0.1, 0.9]^3$ and pairwise correlation,

$$d_{ij} \in \left[\max\left\{-\left(\frac{p_i p_j}{q_i q_j}\right)^{1/2}, -\left(\frac{q_i q_j}{p_i p_j}\right)^{1/2}\right\},\right.$$
$$\left.\min\left\{-\left(\frac{p_i q_j}{q_i p_j}\right)^{1/2}, -\left(\frac{q_i p_j}{p_i q_j}\right)^{1/2}\right\}\right], \ q_i = 1 - p_i, \quad (4)$$

which emit a positive definite Σ matrix (see Emrich and Piedmonte (1991)), were used to generate the data set at each iteration. Similar to before, the performance of each model estimated using a 75% derivation subsample was measured by the SSE values predicting outcomes in the 25% testing subsample. The results comparing the five models,

Table 9. Period 2 validation: predicting joint behaviors by dependence models, joint attraction functions, and copulas

Behavior	*Data*	*IND*	*CIAD*	*CIMD*	*JAF1*	*JAF2*	C_F	C_C
000	773	571	696	702	535	661	687	699
001	2998	2955	2982	3048	2773	2938	3030	3034
010	246	302	324	307	282	297	309	310
011	1372	1563	1390	1334	1459	1320	1362	1346
100	455	521	398	425	436	425	413	403
101	2514	2695	2657	2662	2888	2705	2610	2604
110	194	276	440	252	254	266	259	256
111	1757	1426	1620	1578	1681	1696	1638	1657
SSE	—	235 796	115 480	70 996	265 306	67 911	41 872	36 338

independence, additive dependence, multiplicative dependence, and the two joint attraction functions, over 1000 simulated data sets are displayed in Table 10. Similar to the case of the ARIC health behavior data, we find that the models of conditional dependence perform quite similarly and significantly improve on the assumption of independence. Also, similar to before, we find that JAF1 does not improve a modeler's ability to predict joint behaviors over an assumption of independence. The second joint attraction function also does not improve on the assumption of independence and in fact performs the worst in the simulation study. The result is expected, however, as JAF2 leveraged the insight from the health domain that consistently (un)healthy behaviors occur with increased frequency. Our simulated distributions impose no such relationship on the data. Our interpretation is that joint attraction functions can be used when insights into the process governing the dependence between variables can be identified (e.g., from subject matter experts). This conclusion is important in the healthcare domain, where insights from practitioners, who often have deep experiential knowledge of dependency patterns, can be used to model dependence between health outcomes and behaviors. The copulas also significantly improve on the assumption of independence, on par with the additive and multiplicative dependence models. Given our findings that the dependence models and the copulas perform similarly, we prefer the simplicity of the dependence models, which are parameterized via straightforward statistics and computations, over the copulas' more complicated application and parameterization requiring maximum likelihood optimization.

Table 10. Simulation study: predicting joint behaviors by dependence models, joint attraction functions, and copulas

Model	*IND*	*CIAD*	*CIMD*	*JAF1*	*JAF2*	C_F	C_C
$\overline{SSE}$	182 297	46 899	46 989	249 625	284 684	46 624	51 345
SE_{SSE}	5425	2334	2334	10 563	12 548	2333	2482

6. Conclusions

This article modeled dependencies in health behaviors that affect healthcare costs. These dependencies have implications on benefit designs, policy laws, and regulations involving both the private and public sectors. The main contributions of this article are two general approaches for modeling dependence between multiple behaviors that can be used to predict dependence outside a derivation sample: additive and multiplicative models of conditional probabilities and joint attraction functions. We established a relationship between these two approaches and estimated models of the dependence between diet, exercise, and smoking behaviors. Understanding the dependencies between behavioral risk factors such as these may allow healthcare providers and policy makers to reduce the costs of healthcare and/or improve the quality by better understanding the effects of interventions. For example, since the control of diet, exercise, and smoking behaviors is critical to the prevention and control of chronic diseases, understanding dependencies such as those identified in this article would place an increased benefit on modifying a consumer's exercise behavior. Both of the proposed approaches improved prediction of observed combined exercise, diet, and smoking behavior relative to the more analytically convenient assumption of independence; however, both methods rely on features that may not be present in all application domains. The models of dependence through conditional probabilities in this article leveraged conditional independence observed in the data. Compared to modeling dependence through copulas, the additive and multiplicative dependence models introduced show similar performance in terms of prediction error in both in the ARIC data as well as our simulation study. Given these results, and the relative ease of implementation and interpretation, we believe the additive and multiplicative dependence models presented to be an attractive option for modeling dependence. Although a model of conditional independence will generalize a model of pure independence in any case, the amount of improvement relative to the assumption of independent will naturally depend on the validity of the conditional independence assumption. Another disadvantage of the conditional probability approach relative to the joint attraction approach is the need to constrain parameters to ensure meaningful probabilities. Our simulation study clarified that the ability of the joint attraction functions to model dependence relied on our ability to identify properties (e.g., behavioral consistency) of the dependence process. In larger or more complicated data sets, such insights may be hard to identify, making the specification of an appropriate joint attraction function difficult. Another limitation of the joint attraction approach is the need to specify a stochastic choice function.

Another approach to adjudicating between the proposed methods is to consider the "physical" process in the context of the dependence being modeled. For example, is it more reasonable that consumers' attractions to combined behaviors are *a priori* adjusted according to their consistency or that consumers form probabilities of each separate behavior and then adjust these probabilities based on the outcome of a dominant behavior? In the case of health behaviors, both alternatives seem plausible. The first possibility, modeled by a joint attraction function, is a type of premeditated dependence. For example, the consumer would decide: I want to maximize my health by exercising, smoking will inhibit my lungs from delivering oxygen to my muscles so I will not smoke, and I will need proper nutrition to maximize my muscle function so I will follow a healthy diet. The second type of dependence modeled through conditional probabilities is based on the reality that each separate behavior is not chosen simultaneously and that the sequential nature of the behaviors allows preferences for consistency. For example, the consumer thinks: Now that I have exercised and lost weight and feel better, I need to change my diet to a healthy one to keep the gains I have made. Further research into the cognitive processes of individuals as they balance the long-term risks and rewards of multiple health behaviors from immediate consequences (e.g., higher tobacco taxes and higher premiums for tobacco users versus sustained long-term cessation; increased sense of well-being with exercise versus satisfaction from consuming certain high-fat foods) is needed to elucidate which of these postulates is more accurate for each individual. Confirming that some individuals are more affected by interdependent choices compared to others that have an independent focus would have important implications for public policy on population versus individual initiatives of health behavior incentives; for example, whether to reward the community for going smoke-free versus individual health insurance policy premiums of those who are tobacco-free.

Though motivated by healthcare behaviors, our strategies for dealing with independence could be applied in any domain with dependent uncertainties. The particular insights into behavioral dependencies have implications for reducing costs and improving health relative to diseases for which diet, exercise, and smoking are risk factors. For example, based on our results it may be possible to indirectly modify the risk factors of smoking and diet by changing an individual's exercise behavior. Althought future research is needed to corroborate this claim, the non-independence of these behaviors is clear. Therefore models of dependence such as those presented here should be incorporated into policy and decision-making models.

Funding

This research was funded in part by NSF grant CMMI-0944670. The first author wishes to acknowledge financial support for this work from the U.S. Department of Education through GAANN P200A060241.

This article was prepared using ARIC Research Materials obtained from the NHLBI Biologic Specimen and Data Repository Information Coordinating Center and does not necessarily reflect the opinions or views of the ARIC or the NHLBI.

References

Agresti, A. (2007) *An Introduction to Categorical Data Analysis*, second edition, Wiley, Hoboken, NJ.

Ajzen, I. (1991) The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, **50**(2), 179–211.

Bickel, J. and Smith, J. (2006) Optimal sequential exploration: a binary learning model. *Decision Analysis*, **3**(1), 16–32.

Brown, G. (1951) Iterative solutions of games by fictitious play, in *Activity Analysis of Production and Allocation*, Koopmans, T. (ed), Wiley, New York, NY, pp. 374–376.

Camerer, C. and Ho, T. (1999) Experience-weighted attraction learning in normal form games. *Econometrica*, **67**(4), 827–874.

Chaganty, N. and Joe, H. (2006) Range of correlation matrices for dependent Bernoulli random variables. *Biometrika*, **93**(1), 197–206.

Chessa, A., Dekker, R., van Vliet, B., Steyerberg, E., and Habbema, J. (1999) Correlations in uncertainty analysis for medical decision making: an application to heart-valve replacement. *Medical Decision Making*, **19**(3), 276–286.

Chow, C. and Liu, C. (1968) Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory*, **14**(3), 462–467.

Clemen, R. and Reilly, T. (1999) Correlations and copulas for decision and risk analysis. *Management Science*, **45**(2), 208–224.

Ellison, G. (2006) Bounded rationality in industrial organization, in Blundell, R., Newey, W. and Persson, T. (eds), *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, volume 2, Cambridge University Press, New York, NY, pp. 142–174.

Emrich, L. and Piedmonte, M. (1991), A method for generating high-dimensional multivariate binary variates. *American Statistician*, **45**(4), 302–304.

Ferson, S., Nelsen, R., Hajagos, J., Berleant, D., Zhang, J., Tucker, W., Ginzburg, L. and Oberkampf, W. (2004) Dependence in probabilistic modeling, Dempster–Shafer theory, and probability bounds analysis. Technical report. SAND2004-3072, Sandia National Laboratories, Albuquerque, NM.

Fudenberg, D. and Levine, D. (1998) *The Theory of Learning in Games*, MIT Press, Cambridge, MA.

Gastwirth, J. (1984) Statistical methods for analyzing claims of employment discrimination. *Industrial and Labor Relations Review*, **38**(1), 75–86.

Glasgow, R., Goldstein, M., Ockene, J. and Pronk, N. (2004) Translating what we have learned into practice—principles and hypotheses for interventions addressing multiple behaviors in primary care. *American Journal of Preventive Medicine*, **27**(2), 88–101.

Godin, G. and Kok, G. (1996) The theory of planned behavior: a review of its applications to health-related behaviors. *American Journal of Health Promotion*, **11**(2), 87–98.

Keefer, D. (2004) The underlying event model for approximating probabilistic dependence among binary events. *IEEE Transactions on Engineering Management*, **51**(2), 173–182.

Lehmann, E. (1966) Some concepts of dependence. *Annals of Mathematical Statistics*, **37**(5), 1137–1153.

Machina, M. (2004) Nonexpected utility theory, in *Encyclopedia of Actuarial Science*, Teugels, J. and Bjorn, S. (eds), Wiley, New York, NY, pp. 1173–1179.

McDonald, J.H. (2009) *G-test of Independence*, volume 2. Sparky House Publishing, Baltimore, MD.

Nelsen, R. (2006) *An Introduction to Copulas*, second edition, Springer, New York, NY.

Pratt, M., Macera, C. and Wang, G. (2000) Higher direct medical costs associated with physical inactivity. *Physician and Sportsmedicine*. **28**(10), 63–70.

Roth, A. and Erev, I. (1995) Learning in extensive-form games–experimental data and simple dynamic models in the intermediate-term. *Games and Economic Behavior*, **8**(1), 164–212.

Simon, H. (1955) A behavioral model of rational choice. *Quarterly Journal of Economics*, **69**, 99–118.

Vesely, W., Goldberg, F., Roberts, N. and Haasl, D. (1981) *Fault Tree Handbook*, Nuclear Regulatory Commission, Washington, DC.

Appendix

G-test for independence

The G-test for independence (McDonald, 2009) uses the test statistic $G = \sum_{ij} O_{ij} \times \ln(O_{ij}/\hat{\mathbb{E}}_{ij})$ but otherwise is applied similarly to the χ^2 test. Table A1 shows the results of the G-test applied to the ARIC data from Table 1.

Table A1. G-test for pairwise independence between diet (D), exercise (E), and smoking (S)

Test pair	*Third behavior*	χ^2	*p-Value*
DS	0	1.41	0.235
DS	1	4.35	0.037
DS	C	7.76	0.005
ES	0	35.36	0.000
ES	1	16.75	0.000
ES	C	54.11	0.000
ED	0	5.77	0.016
ED	1	44.11	0.000
ED	C	51.88	0.000

Biographies

Brandon Pope is a Program Manager of Business Analysis at Baylor Scott & White Health. His research interests include incentive mechanisms in distributed systems, healthcare service, and population health delivery and financing. Prior to joining Baylor Scott & White, he was an Associate Research Scientist with the Regenstrief Center for Healthcare Engineering and instructor in the School of Industrial Engineering at Purdue University. He received his Ph.D. from Texas A&M University in Industrial and Systems Engineering and his B.S. in Mathematics from Abilene Christian University.

Abhijit Deshmukh is the James J. Solberg Head and Professor in the School of Industrial Engineering at Purdue University. His research interests are in distributed decision making, system complexity, mechanism design, healthcare policy, and cyberinfrastructure for engineering applications. Prior to joining Purdue, he was the Rockwell International Professor in the Department of Industrial & Systems Engineering and Director of the Institute for Manufacturing Systems at Texas A&M University. He has also served as a Program Director in the National Science Foundation's Engineering Directorate and the Office of Cyberinfrastructure. He received his Ph.D. from the School of Industrial Engineering at Purdue University. He is a Fellow of the Institute of Industrial Engineers

Andrew L Johnson is an Associate Professor and co-director of the Laboratory for Energy-Sustainable Operations in the Department of Industrial and Systems Engineering at Texas A&M University. He obtained his B.S. from the Grado Department of Industrial and Systems Engineering at Virginia Tech and his M.S. and Ph.D. from the H. Milton Stewart School of Industrial and Systems Engineering at Georgia Tech. His research interests include productivity and efficiency measurement, benchmarking, and mechanism design. He has worked on research problems in a variety of industries including healthcare, energy, manufacturing, and logistics. He is an associate editor of *IIE Transactions* and a member of IIE, INFORMS, National Eagle Scout Association, and German Club of Virginia Tech. For more information see his website www.andyjohnson.guru

J. James Rohack, MD, FACC, FACP, is the Chief Health Policy Officer at Baylor Scott & White Health, Director of the Scott & White Center for Healthcare Policy, and Senior Staff Cardiologist, Scott & White Clinic. His research interests are in healthcare delivery and financing at local, state, and national levels and how the intersection of individual choice impacts how delivery and financing needs to be structure in an ethically responsible manner. He is a Professor of Medicine and Humanities at the Texas A&M Health Science Center, where he holds The William R. Courtney Centennial Endowed Chair in Medical Humanitites. He received his B.S. in Psychology from the University of Texas at El Paso and his M.D. from the University of Texas Medical Branch in Galveston. He completed his internal medicine residency, chief residency, and cardiology fellowship at UTMB–Galveston. He is a member of the Texas Medical Association, the American Medical Association, the American College of Physicians, the American College of Cardiology, the American Heart Association, the American College of Physician Executives, and Alpha Omega Alpha.