

COMBINING LIFE TABLE DATA

Gilbert W. Fellingham* and H. Dennis Tolley†

ABSTRACT

Actuaries are often required to form opinions or make value assessments entailing data from several tables, possibly generated from different sources. Often, no formal methods of combining tables are used. This paper illustrates a method for combining tables of count data using maximum likelihood methods. Combining data tables can be problematic for several reasons. Often, not all cells will contain data when multiple tables are combined. Also, since the same level of aggregation over covariates is often not available in each of the constituent tables, some data will exist only on the margins. The method we present is appropriate when data for some cells are missing and even when data are available only on the margins. To illustrate the method, we combine mortality tables from three different sources with different classification information. The analysis indicates the possible presence of a health risk factor beyond the use of alcohol and tobacco in a population of active members of a church.

1. INTRODUCTION

The purpose of this paper is to demonstrate a systematic method of combining data from different tables. Pricing, valuation, and risk assessment related to a block of risk business are usually based on a combination of past experience and expert evaluations of risk. Usually the data representing past experience and the data implicitly being used in expert evaluations can be summarized in tables of counts or histograms. Actuaries are often required, either explicitly or implicitly, to produce the tables from which decisions will be made. Increasingly, these tables are constructed not only from the company's experience, but also from additional sources including related blocks of business, analyses from specialized epidemiological studies, cross-sectional survey data, summaries of experience studies from across the industry, and publicly available government data. Publications by the Society of Actuaries, such as the *Record* (SOA 1996), and other publications such as *Medical Risks* (Lew and Gajewski 1990), derive their value from the fact that actuaries will price product, perform valuations, or make general assessments of risk by explicitly or implicitly

combining these data with other data. *Medical Selection of Life Risks* (Brackenridge and Elder 1992) also contains a variety of information for combining data, but with a more general focus for physicians and underwriters. The recent paper by Brockett et al. (1995) on the use of secondary grouped data is motivated by the need for actuaries to incorporate data other than their own primary experience into the decision-making process.

Although there is a growing collection of data on morbidity and mortality arising from many sources, combining such data has been done either by using ad hoc methods or by fitting each data set to a model and then combining models. In either case, information contained in the original tables is often lost and "information" not in the original tables is created in the combining process. In addition, areas where information is weak or uncertain cannot be easily identified. Assumptions implicit in accepting the resulting models are often hard to state.

Combining information contained in multiple data sets is the focus of a growing literature entitled "meta-analysis" (see, for example, Hedges and Olkin 1985). Meta-analytic methods focus on combining the summary statistics from similar studies to determine the persistence of certain effects over studies. As illustrated in Brockett et al. (1995), in many situations the data are similar in their underlying focus but not in their tabulation or summary statistics. Very little has appeared in the literature on combining tabled data sets beyond meta-analysis. Thus, the additional

*Gilbert W. Fellingham, Ph.D., is an Associate Professor in the Department of Statistics, Brigham Young University, Provo, Utah 84602, e-mail, gwf@byu.edu.

†H. Dennis Tolley, A.S.A., Ph.D., is a Professor in the Department of Statistics, Brigham Young University, Provo, Utah 84602, e-mail, toolley@byu.edu.

information available from using raw data, or even aggregated data giving marginal totals, would provide more extensive conclusions than simply using the summary statistics methods of meta-analysis.

These methods can now be implemented to provide the actuary with additional tools for dealing with multiple sources of data.

We will illustrate a method of combining data from several tables into a single information table. The method illustrated uses maximum likelihood procedures to model the combined data. These likelihood procedures are based on the methodology presented by Tolley and Fellingham (1999) who provide both estimation techniques and address identifiability problems when some of the data combined are aggregated at different levels for some tables relative to other tables. In this paper we show how to implement these procedures in practice.

2. A CONCEPT-FIXING EXAMPLE

To fix ideas, consider the following three sets of data.

1. *National*. This data set was constructed as a synthetic data set, for illustration, using the Illustrative Life Table contained in Bowers et al. (1986). These data are aggregated over sex, smoking status, and alcohol consumption status. Mortality data was available by yearly age groups, and was regrouped into 11 five-year age classifications, starting with ages 30–34 and ending with ages 80–84. After grouping, there were 22 observations from this data set. This data set is shown in Table 1.

Table 1
The National Data by Status and Age

Age	Dead	Alive
30–34	669	480,417
35–39	882	476,519
40–44	1,327	470,884
45–49	2,073	462,157
50–54	3,178	448,594
55–59	4,622	428,537
60–64	6,619	399,669
65–69	8,859	359,914
70–74	11,449	307,888
75–79	13,619	243,967
80–84	15,220	170,447

2. *Church*. This set consists of a ten-year experience study of insured individuals working for organizations associated with a church. The experience was between 1981 and 1991. As a condition of employment, individuals could not be current users of alcohol or tobacco, although they might be former users. Both sex and age information were available. Although age was tabulated in single-year age groups, for simplicity we divided age into five-year classifications. There were 11 such age groups, starting with ages 30–34 and ending with 80–84. The data were predominantly for individuals in the U.S., though there were some employees residing in Canada. There were 44 observations from this data set. This data set is shown in Table 2.

3. *CPSI*. This set consists of the CPSI data set from a study undertaken in the U.S. by the American Cancer Society. This data set contains mortality outcomes for a group of people followed prospectively in time (Lew and Garfinkel 1987). Individuals were enlisted into the study on a voluntary basis with no conditions of representativeness or randomness made. For each individual enlisted, smoking history was determined as well as age and sex. Age was tabulated in nine five-year age groups from 35–39 through 75–79. The study was initiated in 1959 and mortality data are available to 1972. There were 144 observations from this data set, which is shown in Table 3.

The problem considered in this paper is to combine these three mortality data files to form an estimate of

Table 2
The Church Data by Sex, Status, and Age

Age	Female		Male	
	Dead	Alive	Dead	Alive
30–34	2	4,432	8	17,087
35–39	4	4,203	4	16,772
40–44	0	4,711	10	15,075
45–49	3	5,034	7	13,064
50–54	4	5,295	17	11,229
55–59	9	5,572	16	10,280
60–64	11	5,194	29	8,198
65–69	16	3,729	22	4,636
70–74	30	3,364	51	3,923
75–79	44	2,106	91	2,686
80–84	45	1,133	78	1,448

Table 3
The CPS1 Data Set by Smoking History, Sex, Status, and Age

Age	Never Smoker				Former Smoker				Moderate Smoker				Heavy Smoker			
	Female		Male		Female		Male		Female		Male		Female		Male	
	Dead	Alive	Dead	Alive	Dead	Alive	Dead	Alive	Dead	Alive	Dead	Alive	Dead	Alive	Dead	Alive
35-39	70	68,688	19	15,505	8	6,013	6	4,383	63	52,522	53	24,503	14	9,266	40	15,158
40-44	271	179,222	63	33,261	30	16,503	25	10,721	250	127,857	187	51,824	67	21,588	170	33,234
45-49	815	376,039	195	75,801	97	35,089	107	29,712	759	244,665	720	113,267	177	38,592	579	76,613
50-54	1,953	591,937	624	143,779	219	50,654	410	65,231	1,609	335,178	1,980	205,673	332	49,221	1,715	136,451
55-59	3,097	697,346	1,361	183,344	256	49,804	994	89,756	2,203	313,184	3,810	246,225	411	42,328	2,879	152,861
60-64	4,748	650,108	2,102	163,347	297	34,053	1,492	81,378	2,172	202,788	4,731	195,184	386	24,093	3,156	103,561
65-69	6,619	537,215	2,788	126,763	242	19,883	1,659	58,748	1,834	106,890	4,743	123,114	211	10,764	2,395	50,089
70-74	8,374	397,444	3,301	91,714	289	10,799	1,756	35,358	1,397	51,084	4,049	66,339	146	4,000	1,408	18,808
75-79	9,872	248,723	3,633	57,116	234	5,070	1,160	16,405	1,027	21,218	2,750	28,981	83	1,119	621	5,394

mortality for each classification (that is, age, sex, alcohol consumption, and smoking status). We recognize that one of the data sets is artificial (Bowers et al. 1986) and one is dated (Lew and Garfinkel 1987). Additionally, there is no exposure information available in data set one. Therefore, the following example is intended more for illustration of the method than for determining real mortality patterns or related actuarial functions. Since data are available at different levels of aggregation as regards smoking status and alcohol consumption for the various groups, it is necessary to borrow information from some of the mortality studies to apply to the others. In this paper we illustrate how the likelihood solution of Tolley and Fellingham (1999) can be implemented in combining life table data. This solution is based on the Poisson distribution and takes into account constraints imposed both by the sampling method and by the sparseness of data. Since the solution is based on the likelihood principle, estimates of effects that may only be present in small studies will be less precise. However, no information regarding consistent effects will be lost through the combination of the tables. One reason for this, as shown below, is that we retain a parameter identifying each study in the model. On the other hand, suppose that there are several small studies combined, each of which shows a similar pattern. Because the studies are small, the pattern may not be considered important or identified as statistically significant. By combining the tables using the procedures discussed here, the statistical significance of effects persistent across studies is easier to identify.

3. PRELIMINARIES

We assume that each of the data sets to be combined can be put into the format of a multiway contingency table. In such a table, each count is classified into one cell of the table by the characteristics associated with the count. For example, age and mortality outcome are often used to define life table counts, with additional variables such as sex and smoking status defining other classifications. The contingency table template we use will be the one defined by the classification variables available in any of the data sets. The levels of the classification variables will be those that are the least aggregated in any one of the data sets. In the example above, information is available in at least one of the studies on each of the following: (i) smoking status, (ii) alcohol consumption status, (iii) sex, (iv) age, and (v) mortality outcome. Thus, the template we use is a five-way contingency

table. Not all classification variables are available for each data set. For example, data set one has no data on sex, smoking status, or alcohol consumption status. Data set two consists only of currently non-drinking, nonsmoking individuals, who could however be either lifelong nonsmokers or former smokers. In these cases, only the marginal information aggregated across some of the levels of the classification variables is available. Mathematically, we set up the table as if data were present in every cell and then impose constraints to adjust to the data available. This not only helps the analyst understand the constraints inherent in the data, but also aids in forming the component steps required in implementing the statistical methods. In the example we have the following levels represented: (1) age group—11 levels, (2) status—two levels (alive or dead), (3) sex—two levels, (4) smoking—four levels (never, former, moderate, heavy), (5) alcohol—two levels (no, yes), and (6) study—three levels. If data were present in every possible cell, we would have 1,056 cells with data.

To combine data from several tables when some cells have no counts, we must make some assumptions. Imposing assumptions is common. For example, if a table reports mortality rates for 60-year-olds, 65-year-olds, and 70-year-olds, analysts will implicitly fill in the ages between 60 and 70 using some smooth function. Construction of mortality tables from the collective experience of several companies is often smoothed or graduated. The assumptions implicit in the smoothing process allow the analyst to “borrow strength” from ages where data are available to reliably impute rates at other ages.

The methodology illustrated here also uses constraints to smooth and fill in. However, the methods are more formal in that they follow directly from likelihood procedures. As pointed out by a reviewer, the method could also be built on a Bayesian framework. A Bayesian approach would entail some involved integration. However, with current numerical techniques such as the Metropolis algorithm, Gibbs sampling, and other Markov chain Monte Carlo methods, these numerical problems could be resolved (Gilks et al. 1996). Because considerable a priori information is available on morbidity and mortality processes, a Bayesian extension of the current methodology holds much promise.

To develop a framework on which to impose constraints, we define the following. We call \mathbf{c} the vector of (hypothetical) data if there was an observation in every cell. We represent the expected value of this vector as \mathbf{m} ,

$$E[\mathbf{c}] = \mathbf{m}. \tag{3.1}$$

We assume that the components of \mathbf{c} are distributed as Poisson random variables. We also assume that \mathbf{m} can be expressed as

$$\mathbf{m} = e^{\mathbf{X}\boldsymbol{\beta}}, \tag{3.2}$$

where $\boldsymbol{\beta}$ is a vector of unknown parameters and \mathbf{X} is a matrix of known constants. This is the traditional log-linear model formulation of Bishop, Fienberg, and Holland (1975). The notation $e^{\mathbf{X}\boldsymbol{\beta}}$ refers to taking the exponential of each entry in $\mathbf{X}\boldsymbol{\beta}$, component-wise, and placing it in a vector of the same length as $\mathbf{X}\boldsymbol{\beta}$.

Equation (3.2) may be rewritten as

$$\mathbf{X}\boldsymbol{\beta} = \log(\mathbf{m}) \tag{3.3}$$

where $\log(\mathbf{m})$ is a vector obtained by taking the natural logarithm of each entry of \mathbf{m} . In essence, Equation (3.3) says that the unknown parameters $\boldsymbol{\beta}$ are linear combinations of the logarithms of the expected counts. In the simplest case (data present in every cell) the matrix \mathbf{X} is of full rank. In this case, each of the elements in $\boldsymbol{\beta}$ can be explicitly expressed using

$$\boldsymbol{\beta} = \mathbf{X}^{-1} \log(\mathbf{m}), \tag{3.4}$$

where \mathbf{X}^{-1} is the inverse of \mathbf{X} .

In the case of interest, \mathbf{X} is singular. This means that there are some linear combinations of the logarithms of the cell counts \mathbf{m} that are zero. Such an assumption has the effect of smoothing the observed data to estimate \mathbf{m} . In effect, when \mathbf{X} is singular, Equation (3.2) imposes constraints on the model. These constraints are the key to building tables of data when data from different studies have different aggregation patterns.

To illustrate the relationship between imposed assumptions and the data needed, suppose we have a two-by-two table of counts of individuals by smoking classification (yes-no) and alcohol consumption classification (yes-no). A generic table is given in Figure 1. In this figure n_{ij} , $i = 1, 2, j = 1, 2$, are the observed counts of individuals, and n_{i+} and n_{+j} are the row and column totals. We model the expected counts of n_{ij} as Equation (3.1) with the \mathbf{X} matrix given by:

$$\mathbf{X} = \frac{1}{4} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}.$$

Figure 1
Generic Table of Count Data
Classified by Smoking Status
and Alcohol-Consumption Status

		Smoke		
		No	Yes	
Alcohol	No	n_{11}	n_{12}	n_{1+}
	Consumption	Yes	n_{21}	n_{22}
		n_{+1}	n_{+2}	n_{++}

The inverse of this matrix is simply:

$$\mathbf{X}^{-1} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}.$$

This value of \mathbf{X}^{-1} indicates that β_1 , the first entry in $\boldsymbol{\beta}$, is the sum of the entries of $\log(\mathbf{m})$ —that is, by multiplying out in Equation (3.4) we get $\beta_1 = \log(m_1) + \log(m_2) + \log(m_3) + \log(m_4)$. Similarly, $\beta_2 = \log(m_1) + \log(m_2) - \log(m_3) - \log(m_4)$. The fourth term, $\beta_4 = \log(m_1) - \log(m_2) - \log(m_3) + \log(m_4)$, represents an estimate of the interaction between alcohol consumption and smoking. The likelihood for this model conditional on a fixed value for n_{++} is the multinomial likelihood,

$$L = \frac{n_{++}!}{\prod_{i,j} n_{ij}!} \prod_{i,j} \left(\frac{m_{ij}}{N} \right)^{n_{ij}}. \tag{3.5}$$

Taking logarithms and replacing the $\log(m_{ij})$ by the result in Equation (3.3) we get

$$\begin{aligned} \ln(L) = & C + \beta_1(n_{11} + n_{12} + n_{21} + n_{22}) \\ & + \beta_2[n_{11} + n_{12} - (n_{21} + n_{22})] \\ & + \beta_3[n_{11} + n_{21} - (n_{12} + n_{22})] \\ & + \beta_4[n_{11} + n_{22} - (n_{12} + n_{21})], \end{aligned} \tag{3.6}$$

where C is a constant with respect to the β_i 's. Since the multinomial likelihood is a member of the exponential family of probability density functions, the data necessary to obtain the maximum likelihood estimates consist of the data in parentheses following each β coefficient in Equation (3.6) (see Bishop, Fienberg, and Holland 1975). Quick examination confirms that we need each of the n_{ij} values.

Suppose now that we assume $\beta_4 = 0$. Under this assumption, we can remove the last term in the expression of $\ln(L)$ given in Equation (3.6). Again the data necessary to obtain maximum likelihood estimates of the β 's are given in the parentheses following them. Notice that the data necessary consist of the following four sums:

$$n_{11} + n_{12} = n_{1+}$$

$$n_{21} + n_{22} = n_{2+}$$

$$n_{11} + n_{21} = n_{+1}$$

$$n_{12} + n_{22} = n_{+2}$$

Put in other words, given that β_4 is assumed to be zero, only the marginal totals are necessary to obtain the maximum likelihood estimates. We don't need any of the individual data!

This result can be viewed from the opposite direction. Explicitly, if the only data available from the table in Figure 1 are the marginal totals, then all parameters except β_4 can be estimated, provided β_4 is assumed to be zero. This result holds for larger, more complex tables as well. If there are only marginal data at some levels of the table, then only a few of the parameters will not be estimable. However, assuming that these nonestimable parameters are zero results in the ability to form maximum likelihood estimates of the remaining parameters.

The estimation result given above has been identified in the literature and used to solve certain missing data problems. Most applications involve relatively small multiway tables. In the applications below we use more complicated tables with a variety of patterns of marginal sums. However, the identification of which parameters must be set to zero and which can be estimated requires computational steps more mathematically sophisticated than in the illustration above.

To summarize, when the matrix X in Equation (3.3) has more rows than columns (that is, more rows than the column vector β), then the expected cell counts given in Equation (3.2) are implicitly constrained. This means that not all cells need to have data to make estimates of m . To generate these constraints in X , we construct the matrix X^{-1} to be linear combinations of the logarithms of the cell counts that would be of interest assuming all data were present. We then eliminate those columns of X which correspond to the rows of X^{-1} that represent the constraints to be imposed. If these constraints are chosen judiciously, then data in every cell are not necessary to provide smoothed estimates of the expected counts

in each cell. Since each column of X corresponds to a degree of freedom in the fitted model, removal of a column of X by assuming the appropriate constraint also eliminates a degree of freedom in the model.

Exactly how to determine which terms are estimable from the available data and how to implement an estimation procedure are the topics of the rest of this paper. We refer the reader interested in more mathematical detail to the paper by Tolley and Fellingham (1999).

4. THE ANALYSIS MATRICES

We now present in detail a certain matrix notation that is necessary for the analysis. First recall that m , representing the vector of expected cell counts, must have its number of rows equal to the possible number of cells. For our example (see Section 2), m will be of dimension $1,056 \times 1$. Although this vector is hypothetical, it is very important to keep track of the subscripts associated with each element of m . In our example, we used the following notation. Each element of m is designated $m_{i,j,k,l,m,n}$, where the subscripts are as follows:

i	study number	$i = 1(National),$ $2(Church), 3(CPS1)$
j	alcohol use	$j = 1(no), 2(yes)$
k	tobacco use	$k = 1(never), 2(former),$ $3(moderate), 4(heavy)$
l	sex	$l = 1(male), 2(female)$
m	status (alive or dead)	$m = 1(alive), 2(dead)$
n	age group	$n = 1(30-34), 2(35-39),$ $\dots, 11(80-84).$

We assume the subscripts move fastest at the far right (lexicographical order), so that the first 11 expected counts are for study 1, alcohol use 1, tobacco use 1, sex 1, status 1, and age groups 1, . . . , 11. The next 11 observations would be from status group 2, and so on. It is important to keep track of the relative positions of the various cells in the m vector so that other necessary matrices that represent linear combinations of m will be constructed appropriately.

The actual data are kept in a vector which we call n . For these three data sets, the entire data vector n has dimension 210×1 . That is, there are 210 observations available for the analysis. The vector of observations for this analysis was formed by stacking the columns of data in Tables 1, 2, and 3. So the first element of the vector is 669, the second is 882, the twelfth is 480, 417, and so on.

We define a matrix W to identify the cells from which the actual observations were drawn. The i^{th} observed data point will correspond to the i^{th} row of W . The cell (or cells) from which this data point is drawn is denoted by a 1 in the appropriate column of W . For example, a 1 in column 2 of the first row of W would indicate that the first observation came from cell number 2. (This is not true for our example, but entered here only for illustration.) In case the observation is a marginal count—a sum across one or more classification levels—then for the row corresponding to this observation there will be a 1 in each column whose cell value is included in the sum. All other entries of this row will be 0. The W matrix for the example has a row corresponding to each of the 210 observed counts and a column corresponding to each of the 1,056 cells in the complete combined contingency table. The dimensions of the W matrix for this problem are $210 \times 1,056$.

For the example considered here, none of the studies supplied data for individual cells; all data were from margins. Building the W matrix for this particular problem is mainly a bookkeeping issue. Each row of the W matrix consists of 0's and 1's, with a 1 corresponding to each cell which contributed to the total count in the observation represented by the row. In the complete data case, the W matrix would be an identity, representing one observation from each cell. In this problem, each row contained multiple 1's, indicating that each observation represented a sum over a number of different cells. Thus, $W\mathbf{m}$ yields the expected counts of the actual data.

We now construct the first row of the W matrix explicitly. The first row of the W matrix must correspond to the first data element of the vector \mathbf{n} . This element is the 669 count of subjects in the National study whose status is “dead” and whose age is 30–34. These subjects comprised both sexes from all alcohol- and tobacco-use levels. We now recall how the \mathbf{m} vector was constructed. The first one-third (352 of the 1,056 cells) of the elements of \mathbf{m} represent the National study. Elements 1, 12, 23, . . . represent ages 30–34, elements 2, 13, 24, . . . represent ages 35–39, and so on. Elements 1–11, 23–33, 45–55, . . . represent “alive” while elements 12–22, 34–44, 56–66, . . . represent “dead.” Sex would be represented in blocks of 22, with the first 22 cells representing males, the next 22 females, the next 22 males, and so on. Tobacco use is in blocks of 44, with the first 44 elements for “never,” the next 44 for “former,” the next 44 for “moderate,” and the next 44 for “heavy.” That cycle repeats six times. Finally, non-alcohol users

would be represented in the first 176 cells and alcohol users in the next 176 cells, with this cycle repeated three times. With this in mind, the first observation of 669 would include people from cell 12 (age 30–34, status dead, sex male, tobacco never, alcohol no, and study 1), cell 34 (age 30–34, status dead, sex female, tobacco never, alcohol no, and study 1), cell 56 (age 30–34, status dead, sex male, tobacco moderate, alcohol no, and study 1), and so on. Thus, the first row of the W matrix would consist of 1's in cells 12, 34, 56, . . . , 342, and 0's everywhere else. The next 209 rows of the W matrix are produced analogously. Any programming language capable of handling matrices may be used to construct this matrix. The code is simplified if the language allows both matrix manipulation and manipulation of matrix elements.

To construct the X^{-1} matrix, recall that this matrix is used to construct linear combinations of the logs of the expected cell counts (\mathbf{m}). Thus, we must keep track of the elements of \mathbf{m} as we construct X^{-1} the same way we did to construct W . These linear combinations will, we hope, be those of most interest to the researcher. However, we need to keep in mind that because of data sparseness, not all linear combinations will be estimable.

Initially we work as if data were present in all cells. We construct our X^{-1} matrix as follows. The first row computes the overall sum of the log cell counts. This is simply a row of 1's that yields the sum of the log cell counts when multiplied by $\log(\mathbf{m})$. The next ten rows are orthogonal polynomials constructed to estimate linear, quadratic, cubic, and so on, up to the 10th degree functions of age. Though one may select other combinations, the polynomial orders are appealing. We know that with the sparseness of the data many combinations are not estimable. If we had to forfeit high-order polynomials of the age effect and approximate the age effect as, say, a fifth-order polynomial, we would expect little loss in this assumption. Orthogonal polynomials may be automatically constructed using a number of computer packages. We used the function `orpol` in SAS Proc IML (SAS Institute 1990). Every 12th cell the coefficients will start to repeat, as there are 11 age groups. These coefficients will be repeated in 96 blocks of 11 along the rows of the X^{-1} matrix because age moves the fastest in the \mathbf{m} vector. The next row, 12, computes the difference between the number of alive and the number of dead, or what we called the status effect. Because the \mathbf{m} vector has 11 cells from status = alive followed by 11 cells from status = dead, the row of the X^{-1}

matrix will have 48 groups of 11 -1 's followed by 11 1 's.

Row 13 estimates the sex effect. Because there are 22 cells associated with males followed by 22 cells associated with females in the \mathbf{m} vector, this row of the \mathbf{X}^{-1} matrix will be 24 groups of 22 -1 's followed by 22 1 's. Here we model the difference in sex effect. One advantage of this over simply computing the "additional" effect of female is that the resulting column maintains an orthogonality that results in a slight pedagogical advantage.

The next three rows (14 through 16) were constructed to estimate the effect of smoking. The first contrasted level one (never smoked) with level two (former smoker) and was constructed as 6 groups of 44 -1 's followed by 44 1 's followed by 88 0 's. The next row contrasted level three (moderate smoker) with level four (heavy smoker), and also had 6 groups, this time with 88 0 's followed by 44 -1 's followed by 44 1 's. The final degree of freedom for smoking contrasted levels one and two (never and former smokers) with levels three and four (moderate and heavy smokers). This row consisted of 6 groups of 88 -1 's followed by 88 1 's.

The next row estimated the alcohol effect. As with sex we model the difference. This row consisted of 3 groups of 176 -1 's followed by 176 1 's. The next two degrees of freedom were associated with the effect of study. Since the three data sets were gathered a number of years apart, these linear combinations were constructed to look for a time effect. The first of these contrasted the Church study (the most recent) against the CPS1 study (the earliest). This row of the \mathbf{X}^{-1} matrix consisted of 352 0 's followed by 352 1 's followed by 352 -1 's. The second degree of freedom for study was a quadratic effect of time, the Church, and CPS1 data contrasted against the National data (collected between the other two). This row consisted of 352 -2 's followed by 704 1 's. These were all the degrees of freedom for the main effects. The other degrees of freedom were constructed as interactions among these main effects.

The interaction rows of the \mathbf{X}^{-1} matrix are calculated from the main effects just constructed as componentwise products of the entries of these previous rows. Higher order interactions are three-way, four-way, and so on products of the components of the main effect columns. For example, the interaction between the linear age effect and sex is calculated by forming the row of products of the individual components in the linear age row of \mathbf{X}^{-1} , row 2, with the corresponding entry in row 14 (sex) and placing the

result in the row corresponding to the interaction term. Three-way and higher interaction terms are produced analogously. Since \mathbf{X}^{-1} must be square and of the same row dimension as \mathbf{m} , the dimensions of the \mathbf{X}^{-1} matrix for our example are $1,056 \times 1,056$.

The final matrix that needs to be understood is the \mathbf{Z} matrix. This matrix describes which of the observed totals are constrained by sampling considerations. Not all of the counts vary stochastically—some are fixed. For example, the number of dead plus alive at a certain age is a fixed total, so these totals are constrained. The \mathbf{Z} matrix is again mostly zeros, with ones in each row identifying which cell counts must total to a fixed number. Since the total over the term status was fixed, a row of the \mathbf{Z} matrix will have a 1 in the column associated with "alive" and a 1 in the column associated with "dead" for a given pair of observations.

The \mathbf{Z} matrix only operates on observed data, so as we build the \mathbf{Z} matrix, we need only concern ourselves with the form of the observed data vector \mathbf{n} . The first number in \mathbf{n} is 669, and the twelfth number is 480,417. Since these numbers represent the total number of subjects in the 30–34 age group for the National study, their sum is a constrained total. Thus, the first row of the \mathbf{Z} matrix will contain a 1 in the first and twelfth positions and 0 's elsewhere. The rest of the rows of \mathbf{Z} are produced analogously. There are 210 observed data points with 105 data pairs, so the dimensions of \mathbf{Z} are 105×210 .

5. BUILDING THE MODEL

Because the only data available are the marginal counts represented by \mathbf{n} , and because we are subject to those constraints made explicit in the \mathbf{Z} matrix, we are restricted as to which β 's are actually estimable. In fact, the likelihood function itself (see below) ultimately determines which of the β 's may actually be included in the final model.

In this analysis, since alive-or-dead status is a constrained total, we really have a bivariate measure in each of 528 cells. That is, if we think of the cells as the proportion of alive and dead in each setting, we are constrained that those proportions add up to 1 . We are limited by that constraint to focus on the degree of freedom associated with the main effect for status and all interactions involving status. These account for the 528 degrees of freedom that would be estimable if all data were present.

For the sake of simplicity, we also limited our search for estimable degrees of freedom to three-way

interactions or lower and to polynomials of up to only the fifth degree involving age. This left our possible search space at 65 degrees of freedom. These degrees of freedom include status (the purpose of this term is to standardize the response, much like an intercept term in a traditional ANOVA), status by study1, status by study2, status by alcohol, status by smoking1, status by smoking2, status by smoking3, status by sex, and status by age-linear through status by age-quintic (these terms are essentially the “main effects” of the analysis). The remaining terms are all three-way interactions involving status and the “main effects” previously mentioned. Since these are the only degrees of freedom we wish to examine, we define the matrix \mathbf{X} to have 65 columns. These 65 columns correspond to the 65 rows of \mathbf{X}^{-1} which we wish to estimate. The remaining rows of \mathbf{X}^{-1} therefore must define linear combinations of the log cell counts, which are constrained to be 0. Thus, the \mathbf{X} matrix will be of dimension $1,056 \times 65$.

However, not even all of these 65 degrees of freedom will be estimable. A final determination of estimability is made by appealing to the Implicit Function Theorem (Apostol 1957). As shown in Tolley and Fellingham (1999), this results in assessing the rank of a matrix whose individual components are formed from the derivatives of the likelihood. Since we use Newton-Raphson as our estimation procedure, this matrix must be computed. Thus, if any of the degrees of freedom (terms of the model) are not estimable, the Jacobian will not be invertible, and the computation will fail. We could identify the β parameters corresponding to columns of \mathbf{X} by performing a Gram-Schmidt orthogonalization of the Jacobian. Those rows “zeroing out” correspond to the columns of \mathbf{X} that must be removed. These in turn correspond to the entries of β that must be assumed to be 0. Alternatively, we may start with a “small” model, meaning a model with few terms. This model is checked to make sure it is estimable by checking the Jacobian. We would then add terms to the model one at a time, checking at each step for estimability. This second procedure was the one we used for the example.

When using the Newton-Raphson method to estimate the β 's, it is important to get initial estimates within the radius of convergence. This is not always a

trivial matter. We are currently exploring numerical options that may make this step easier to implement. We built our models by first finding estimates for a small model, which had essentially main effects (the two-way interactions involving status are main effects in this model), and then, using these estimates for the starting point, testing gradually more complicated models. We used $[(\mathbf{WX})'\mathbf{WX}]^{-1}(\mathbf{WX})'\log(\mathbf{n})$ for the initial estimates, where \mathbf{X} was limited to columns corresponding to the effects for status, status by alcohol consumption, status by tobacco, status by sex, and status by age-linear.

The actual implementation of Newton-Raphson is accomplished by first calculating the expected values of \mathbf{m} from the initial estimates for β (call these \mathbf{b}_0) and the columns of \mathbf{X} , which correspond to the effects you are trying to estimate. These estimated values are simply

$$\mathbf{m} = e^{\mathbf{X}\mathbf{b}_0}.$$

Next, the first derivative vector of the log-likelihood (call it $\mathbf{1}$) must be calculated. Let $\mathbf{j}210$ be a column vector of 1's of length 210 (the number of rows of \mathbf{W}), let $\mathbf{j}r105$ be a row vector and $\mathbf{j}c105$ be a column vector of 1's of length 105 (the number of rows of \mathbf{Z}), let $\#$ indicate element-by-element matrix multiplication, let $/$ indicate element-by-element matrix division, let $\#\#$ indicate element-by-element raising to a power, and let $*$ indicate standard matrix multiplication. When performing element-by-element operations with a matrix and a vector (for example, $\mathbf{X}\#\mathbf{m}$), duplicate the vector columnwise so that the dimensions are the same. The calculation (see also Tolley and Fellingham 1999) is then

$$\begin{aligned} \mathbf{1} = & [\mathbf{n}/(\mathbf{W} * \mathbf{m})]'\ * [\mathbf{W} * (\mathbf{X}\#\mathbf{m})] \\ & - \mathbf{j}210 * [\mathbf{W} * (\mathbf{X}\#\mathbf{m})] \\ & - \{(\mathbf{Z} * \mathbf{n})/[\mathbf{Z} * (\mathbf{W} * \mathbf{m})]\}' \\ & * \{\mathbf{Z} * [\mathbf{W} * (\mathbf{X}\#\mathbf{m})]\} \\ & + \mathbf{j}r105 * \{\mathbf{Z} * [\mathbf{W} * (\mathbf{X}\#\mathbf{m})]\}. \end{aligned} \quad (5.7)$$

Also, the second derivative matrix of the log-likelihood (call it \mathbf{J}) is calculated as follows:

$$\begin{aligned}
 J = & - [W * (X\#m)]' * \{[n/(W * m)\#\#2] \\
 & \#[W * (X\#m)]\} \\
 & + ((W * X)' * \{n/(W * m)\#[W * (X\#m)]\}) \\
 & - \{(W * X)' * [W * (X\#m)]\} \\
 & + \{Z * [W * (X\#m)]\}' \\
 & * (((Z * n)/\{[Z * (W * m)]\#\#2\}) \\
 & \#[Z * [W * (X\#m)]\}) \\
 & - (((Z' * \{[Z * n]/[Z * (W * m)]\})\#[W * (X\#m)] \\
 & * [W * (X\#m)]) \\
 & + \{[(Z' * j_{c105})\#[W * (X\#m)]\}' * [W * (X\#m)]\}.
 \end{aligned}
 \tag{5.8}$$

Finally, the last step of the Newton-Raphson procedure is to obtain an updated estimate of β , say b_1 , as

$$b_1 = b_0 - J^{-1} * 1'. \tag{5.9}$$

Using the estimate, return to the top of the loop and iterate until convergence.

Once the process has converged, a mortality table can be built as follows. Take the estimates for the β 's (b) and compute expected cell counts $m = \exp(X * b)$, where X is the design matrix from above. For any particular treatment combination (for example, non-smoking, nondrinking males ages 31–34) there are two expected counts, one for alive and one for dead. Dividing the expected count for number dead by the total of alive plus dead gives the expected probability of death for that particular treatment combination. Then, starting with a radix of, say, 100,000 in the youngest age group, these expected probabilities can be used to calculate the number that die during the first age group by multiplication of the probability of death. Decrementing the total alive for the second age

group by this amount, we multiply the result by the probability of death in the second year, and so forth.

6. RESULTS

The first model that we examine has 11 degrees of freedom that are significant; none of the other possible terms are significant. We show the coefficients, their standard errors, chi-square, and p -values in Table 4.

One can see that the effect for alcohol is quite large, perhaps larger than would be expected. The alcohol effect is six times larger than the largest smoking degree of freedom and an order of magnitude larger than the other two smoking degrees of freedom. This result was not anticipated, but further analysis gives some insight into a possible reason.

Figure 2 shows the actual mortality experience of Church males compared with the expected experience for nonsmoking (former and never), nondrinking males. The fit appears to be appropriate.

However, if we look at the expected experience of former-smoker females (both alcohol categories) and compare it to the CPS1 data for former-smoker females, we see that the effect of alcohol, especially in the younger age groups, appears to be overestimated. This is shown in Figure 3. Recall that the observed data are the aggregate of the two alcohol states. The observed data are a weighted average of these two predicted survival curves. The alcohol “effect” is measured as the distance between the two predicted curves. This same general pattern can be seen in all the CPS1 data when compared to expected results.

In Figure 4 we also look at the expected experience of females who have never smoked (the best mortality experience) and heavy-smoker males (the worst mortality experience) along with the actual National data.

Table 4
Coefficients, Standard Errors, and p -Values for Original Fit of Mortality Data

Term	$\hat{\beta}$	StErr	χ^2	p -Value
status	-4.59	0.00540	721,000	0.00
status \times alcohol	0.00275	0.000141	380	0.00
status \times smoking (none vs. former)	0.000180	0.00001	320	0.00
status \times smoking (moderate vs. heavy)	0.000256	0.000009	758	0.00
status \times smoking (none, former vs. mod, heavy)	0.000450	0.000008	3,494	0.00
status \times sex	-0.000588	0.000006	8,563	0.00
status \times age (linear)	0.411	0.00206	39,600	0.00
status \times age (quadratic)	0.0171	0.00164	109	0.00
status \times age (cubic)	-0.0116	0.00115	102	0.00
status \times age (quintic)	-0.00701	0.000769	83.2	0.00
status \times alcohol \times age (linear)	-0.000305	0.0000371	67.9	0.00

Table 5
Coefficients, Standard Errors, and p -Values for the Second Fit of Mortality Data

Term	$\hat{\beta}$	StErr	χ^2	p -Value
status	-4.61	0.00506	830,000	0.00
status \times smoking (none vs. former)	0.000163	0.000011	238	0.00
status \times smoking (moderate vs. heavy)	0.000263	0.000009	784	0.00
status \times smoking (never, former vs. mod, heavy)	0.000473	0.000007	4,150	0.00
status \times sex	-0.000578	0.000006	8,180	0.00
status \times age (linear)	0.418	0.00193	46,800	0.00
status \times age (quadratic)	0.0123	0.00161	58.6	0.00
status \times age (cubic)	-0.0102	0.0011	81.3	0.00
status \times age (quintic)	-0.00746	0.00077	94.8	0.00
status \times alcohol \times age (linear)	0.000218	0.000016	185	0.00
status \times alcohol \times age (quadratic)	-0.000177	0.000017	108	0.00
status \times alcohol \times age (cubic)	-0.000033	0.000017	3.91	0.048

Again we see that the effect of alcohol seems to be overestimated.

Since the alcohol effect appeared to be too large, we ran models without the status \times alcohol term to see what resulted. In Table 5 we show the result from the best fitting model without the alcohol effect in the model.

If we now look at plots of expected rates along with actual data we see an interesting result. In Figure 5 we show a plot of Church males and expected rates. Now we see the expected rates are clearly too high, presumably because of the lack of the alcohol term in the model.

On the other hand, the plot of CPS1 data in Figure 6 shows a much better fit of actual and expected data. This result was repeated for CPS1 data in all subsets of the population.

The fit for the National data, shown in Figure 7, appears to be marginally better as well.

It appears to us that the Church mortality experience is better than would be expected from a lifestyle free from alcohol and tobacco based on the other data sets. Thus, it is difficult to fit all data with only the given risk factors. The procedure is overestimating the effect of alcohol in an attempt to fit the Church data, which are the only source of information on alcohol. Neither of the other data sets has any information on alcohol use. It certainly appears from the data that there is another unmeasured risk factor beyond the use of alcohol and tobacco that is present in the CPS1 and National data that is not present in the Church data.

Figure 2
Expected Mortality Experience of Nonsmoking (Former and Never), Nondrinking Males and Actual Church Rates (Original Fit)

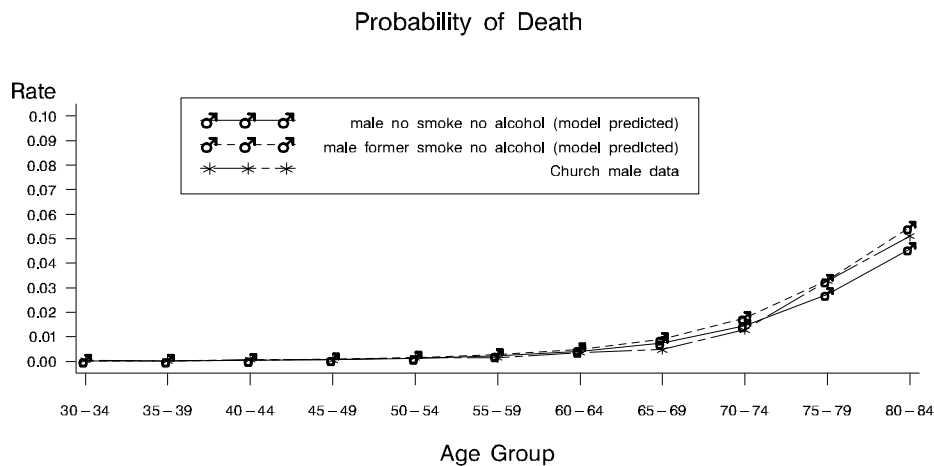


Figure 3
Expected Mortality Experience of Former-Smoker Females
and Actual CPS1 Rates (Original Fit)

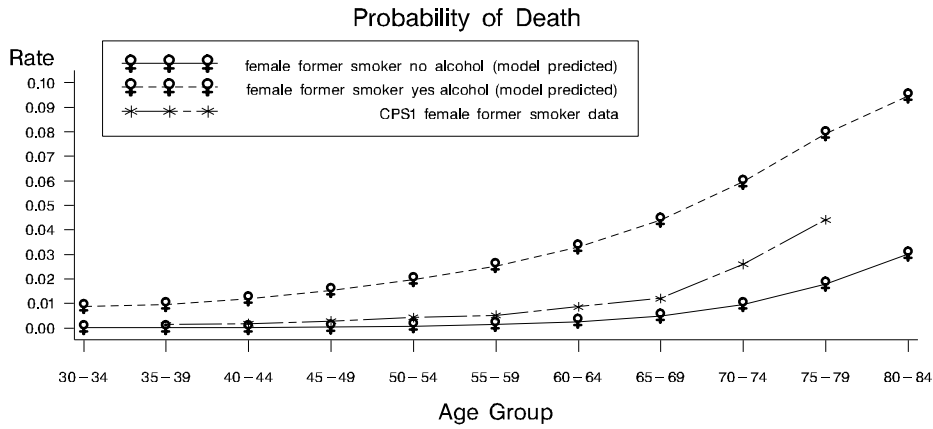


Figure 4
Expected Mortality Experience of Females Who Have Never Smoked and Heavy-Smoker Males
and Actual National Rates (Original Fit)

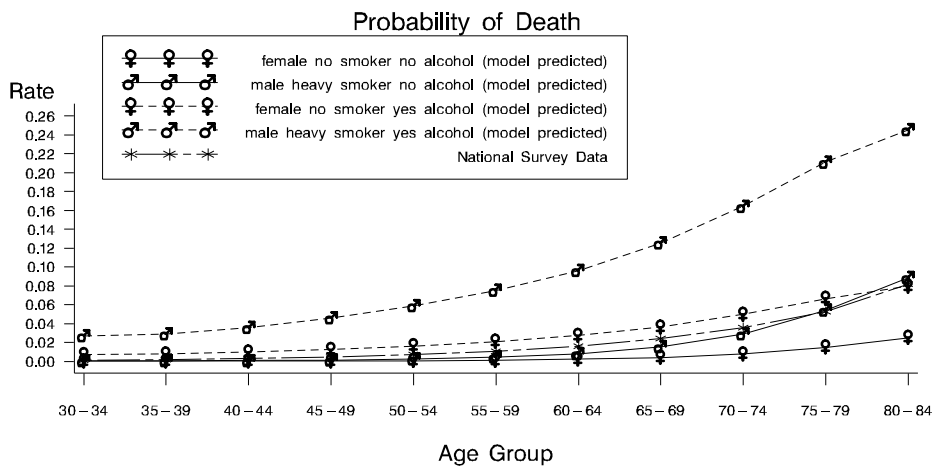


Figure 5
Expected Mortality Experience of Nonsmoking (Former and Never), Nondrinking Males
and Actual Church Rates (Second Fit)

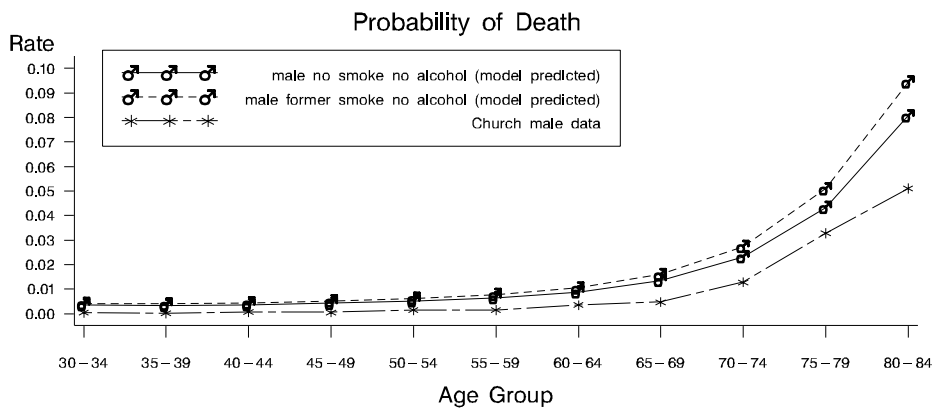


Figure 6
Expected Mortality Experience of Female Former Smokers
and Actual CPS1 Rates (Second Fit)

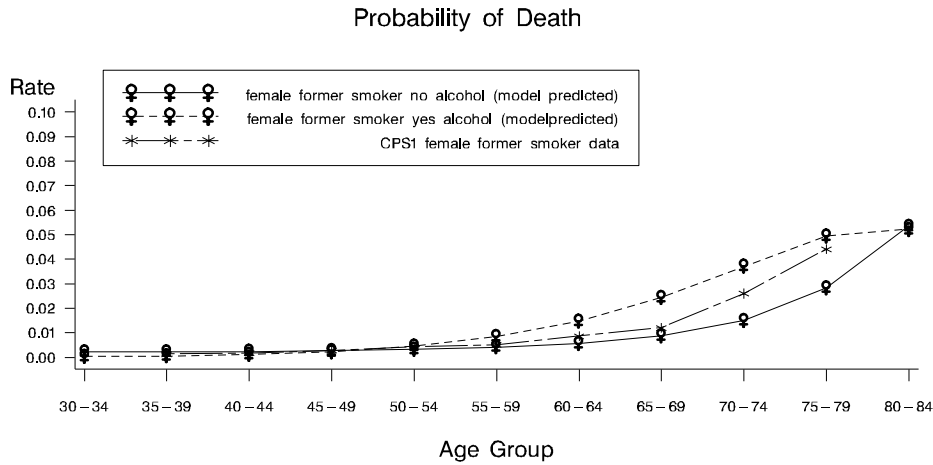
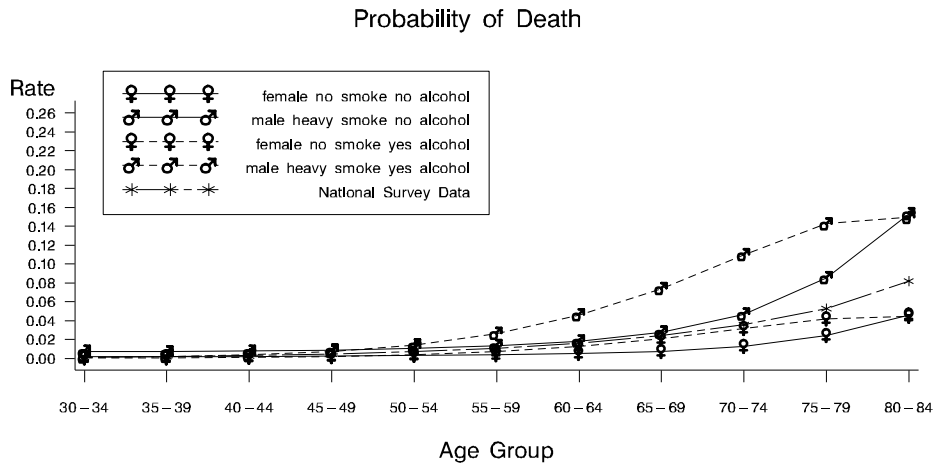


Figure 7
Expected Mortality Experience of Females Who Have Never Smoked
and Heavy-Smoker Males and Actual National Rates (Second Fit)



7. CONCLUSION

We have illustrated a maximum-likelihood technique which allows estimation of effects in log-linear models of categorical data even when many cells are missing data and some data are available only on the margins. We have demonstrated the methodology using three data sets having only marginal information, so none of the available data represents a count in a single cell. The analysis demonstrates the possible presence of a health risk factor beyond the use of alcohol and tobacco in the general population that seems not to be present in the Church population.

REFERENCES

APOSTOL, T. 1957. *Mathematical Analysis*. Reading, Mass.: Addison-Wesley Publishing Co., Inc.

BISHOP, Y.M., FIENBERG, S.E., AND HOLLAND, P.W. 1975. *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, Mass.: The MIT Press.

BOWERS, N.L., GERBER, H.U., HICKMAN, J.C., JONES, D.A., AND NESBITT, C.J. 1986. *Actuarial Mathematics*. Itasca, Ill.: Society of Actuaries.

BRACKENRIDGE, R.D.C., AND ELDER, J.W. 1992. *Medical Selection of Life Risks* (Third edition). New York: Stockton Press.

BROCKETT, P.L., COX, P.L., GOLANG, B., PHILLIPS, F.Y., AND SONG, Y. 1995. "Actuarial Usage of Grouped Data: An Approach

- to Incorporating Secondary Data," *Transactions: Society of Actuaries* XLVII:89–113.
- GILKS, W.R., RICHARDSON, S., AND SPIEGELHALTER, D.J. (eds.) 1996. *Markov Chain Monte Carlo in Practice* (First edition). London, UK: Chapman & Hall.
- HEDGES, L.V., AND OLKIN, I. 1985. *Statistical Methods for Meta-Analysis*. Boston: Academic Press, Inc.
- LEW, E.A., AND GAJEWSKI, J. (eds.) 1990. *Medical Risks: Trends in Mortality by Age and Time Elapsed*. New York: Praeger.
- LEW, E.A., AND GARFINKEL, L. 1987. "Differences in Mortality and Longevity by Sex, Smoking Habits and Health Status" (with discussion). *Transactions of the Society of Actuaries* XXXIX:107–130.
- SAS INSTITUTE, INC. 1990. *SAS/IML Software: Usage and Reference, Version 6* (first edition). Cary, N.C.: SAS Institute, Inc.
- SOCIETY OF ACTUARIES. 1996. *Record*. Schaumburg, Ill.: Society of Actuaries.
- TOLLEY, H.D., AND FELLINGHAM, G.W. 1999. "Likelihood Methods for Combining Tables of Data," *Scandinavian Actuarial Journal* (submitted).

DISCUSSIONS

ERIC STALLARD*

The authors are to be congratulated for writing an excellent paper that provides a comprehensive treatment of an important problem that the actuarial and statistical literature tends to ignore: combining raw data from multiple sources to develop improved estimates of vital rates.

Through the educational syllabus of the Society of Actuaries, most practicing actuaries will have already mastered the basic techniques of rate estimation and graduation (or smoothing) and will have studied applications of credibility theory to situations where existing experience is not large enough to support independent rate estimation. The current paper adds to the repertoire of techniques that can be used by the practicing actuary in dealing with data-analytic situations where the totality of the data is credible, but the data are arranged in the form of multiple subtables that share many—but not all—stratifying dimensions. I have comments in two areas: the presentation of the model and the illustrative application.

The method is based on an application of maximum likelihood principles to log-linear modeling of contingency tables. The fundamental assumption is that the

cell frequencies are Poisson-distributed random counts. The likelihood expression derives from imposition of constraints on certain sums of these counts (encoded in the matrix \mathbf{Z}). The basic model structure derives from the recognition that the expected cell count vector, \mathbf{m} , in Equation (3.2), can be written as a function of a more compact parameter vector, $\boldsymbol{\beta}$, that can be manipulated by the actuary to represent a broad range of effects. For example, the standard approach to hierarchical modeling suggests inclusion of an overall effect, main effects for each stratifying dimension, and any desired set of second- and higher-order interaction terms. These effects can all be encoded in the matrix \mathbf{X} . The contribution of the present paper is that it demonstrates that the imposition of constraints on certain appropriately selected elements of the vector \mathbf{m} is equivalent to the elimination of corresponding columns of the matrix \mathbf{X} . The precise nature of the constraints that must be imposed depends on the detailed characteristics of the component subtables and the stratifying dimensions that are omitted from each.

The steps in model building are relatively straightforward and could be implemented for many applications in about a day's time. The fundamental tasks are:

- Select the data tables to be combined
- Perform any recodes necessary to obtain comparable categorical definitions
- Fix the order of the dimensions in the contingency table template
- Define a mapping from the contingency table template to the vector \mathbf{m}
- Define the constraints matrices \mathbf{Z} and \mathbf{W}
- Define the design matrix \mathbf{X} .

Estimation is conducted via the Newton-Raphson method using a sequence of models of increasing complexity, with each model specification dependent on which specific columns of the matrix \mathbf{X} are eliminated. Expressions are given for the first- and second-order derivatives of the log-likelihood function, but not for the log-likelihood function itself. This latter expression may be useful in situations where there are problems involving convergence of the parameters, as well as in calculating the standard large-sample approximation to chi-square (that is, -2 times the difference in the log-likelihood statistics for two nested models). Expressions are given for obtaining initial estimates of the parameter vector $\boldsymbol{\beta}$, and these may be helpful, especially early in an analysis where one is working with "small" models.

*Eric Stallard, A.S.A., is Research Professor of Demographic Studies at the Center for Demographic Studies, Duke University, 2117 Campus Drive, Durham, North Carolina 27708-0408, e-mail, eric@cds.duke.edu.

Overall, I found the presentation of the model to be quite thorough and the level of detail sufficient to ensure that an actuary attempting to implement this approach could do so with confidence in the accuracy of the results.

As enthusiastic as I was about the development of this new methodology, I was disappointed in the numerical example selected for illustration. Both of the component data sets have characteristics that suggest they are less than desirable for the purposes for which they are used. For example, the national mortality data are not “real” data, but instead are estimates based on a published life table. The CPS1 data are real, but they are not random. Instead, individuals were enlisted into the study on a voluntary basis, and it is possible that the mortality experience of this cohort was subject to selection bias. It is also possible that the smoking composition of this cohort was not representative of the population from which the volunteers were drawn. The church data appear to be the best of the three in that they represent the complete experience of an insured population observed over a ten-year period. Nonetheless, these data are not without difficulty in that the alcohol effect estimated therefrom appears to be about five times larger than the largest smoking effect (see Table 4). This appears to represent some type of confounding of the absence of alcohol use among church members with other characteristics of these same persons that make them less susceptible to mortality.

In general, the analysis of risk factors thought to be predictive of mortality needs to be conducted carefully. When death is classified by cause, it is possible to make more precise estimates of the impact of various risk factors. For example, cigarette smoking has been linked with up to 14 different causes of death in various Surgeon General reports. Alcohol has been linked with heart disease, liver disease, and certain cancers, as well as accidents and violence. However, epidemiologic studies have recently shown that modest alcohol use may in fact lead to a reduction in mortality rates. In attempting to analyze such data, it would be important to take account of the difference between lifelong non-drinkers and nonsmokers and persons who have recently become non-drinkers or nonsmokers due to deteriorating health. Such data are available from some of the larger community-based longitudinal epidemiologic studies, such as the Framingham Heart Study. Alternatively, such data may also be available in proprietary databases maintained by health and life insurance companies. These

latter data are likely to be accessible to practicing actuaries and it would be of interest to see the methods in this paper applied to data sets of this type.

AARON TENENBEIN*

This paper makes an important contribution to the literature on estimating mortality rates. It demonstrates the statistical methodology required to combine mortality experience from different studies in order to obtain overall estimates of mortality rates for various categories of lives. The theoretical development of this statistical methodology is presented in Tolley, Fellingham, and Scott (1998).

The combination of mortality experience from different studies is a very important problem, as it provides estimates of mortality rates for multiple classifications of lives. Single-mortality studies cannot accomplish this task because of limited data or lack of data in particular categories. For example, combining aggregate studies of mortality by sex and age groups, along with three other mortality studies linking smoking status, alcohol usage, and occupational categories with mortality experience, can provide estimated mortality rates for multiple categories based on age, sex, smoking status, alcohol usage, and occupational category.

This problem is important in many other areas. A technique known as “meta-analysis” has been developed to treat this class of problems [Hedges and Olkin (1985)]. Meta-analysis techniques generally combine the results of different studies to perform tests of significance that take into account all of the results of the individual studies. This results in a more powerful test overall, which implies a greater probability of detecting differences that exist in the population. In addition, estimates resulting from combining the different studies are more accurate—that is, they have a smaller sampling error than any of the estimates in the individual studies.

Meta-analysis has been very important in medical research in which studies conducted under different conditions in different localities are combined to measure the efficacy of medical treatments for different diseases. With the explosion in the availability of data from such sources as supermarket scanners and the

*Aaron Tenenbein, A.S.A., is Professor of Statistics and Actuarial Science at the Leonard Stern School of Business, New York University, 44 West Fourth Street, New York, NY 10012, e-mail, atenenbe@stern.nyu.edu.

internet, the use of meta-analysis techniques is becoming important in many areas, such as marketing and consumer preference measurement.

The authors present a meta-analytic approach to combining mortality experience from different studies. The importance of their work is that this combination is carried out using statistical methods based on maximum-likelihood estimation. They also use smoothing techniques to obtain mortality rates at specific ages. Brockett et al. (1995), in a paper mentioned by the authors, treat a similar problem using information theory. These methodologies are much more objective, and more statistically sound, than combining the data by subjective weighting methods.

The example that the authors use is particularly illustrative of the advantages of the maximum-likelihood approach in combining data from different mortality studies. It also demonstrates some of the problems associated with combining data.

In particular, the three data sources (National, Church, and CPS1) allow for the estimation of mortality rates on the basis of age, sex, smoking status, and alcohol usage. The information on smoking status is obtained from the Church and CPS1 data source. The information on alcohol usage is obtained from the Church data. The model can then estimate, for example, the mortality rate of a 42-year-old male who uses alcohol and is a former smoker. The combination of data sources provides a larger sample with more categories, which will produce more precise estimates of the mortality rates.

The three data sources represent mortality experience from different subpopulations. As a result, problems combining data become an issue because mortality experience may be quite different within each subpopulation. The authors demonstrate this problem in their analysis of measuring the effect of alcohol usage. One of the problems is that no study directly measures the mortality rates for a population that is in the alcohol-use category. The Church data represents the population that is in the no-alcohol-use

category, whereas the other data sources represent aggregate data with respect to alcohol usage. Hence, the estimated mortality rates for the alcohol-use population have to be extrapolated. This problem does not occur for tobacco usage, because the CPS1 data contains mortality experience for all of the smoking categories by age and sex. The authors illustrate this point and conclude that the alcohol effect is overstated, because this effect is extrapolated from the Church mortality experience, which is quite different from the experience in the other studies.

This latter problem of combining populations with different mortality characteristics is a very difficult one. One approach to solving this problem may be to weight the different studies before applying smoothing and maximum-likelihood techniques. The weights may be determined by using the proportion of that subpopulation which is present in the entire population. These proportions could be determined from demographic statistics within the framework of a Bayesian analysis. A Bayesian approach was used by Kimmeldorf and Jones (TSA XIX) and Hickman and Miller (TSA XXIX) in their papers on Bayesian graduation. In their work, the Bayesian analysis used a prior distribution based on estimated mortality rates from earlier studies and combined this prior distribution with data on current mortality rates.

In summary, the authors have presented and demonstrated statistical methodology for combining data from different mortality studies to produce estimates of mortality rates for multiple classifications of lives. They clearly demonstrate both the advantages and the problems associated with this technique. Hopefully, this work will continue.

Additional discussions on this paper can be submitted until January 1, 2000. The authors reserve the right to reply to any discussion. See the Submission Guidelines for Authors on the inside back cover for detailed instructions on the submission of discussions.