Evaluation of the Economic Impact of California' s Tobacco Control Program : A Dynamic Model Approach-- Appendix 4 : The Evaluation of TCP through Simulations with Computer Experimental Design.

Leonard S. Miller

Background

We have estimates of the age specific smoking initiation and smoking cessation rates in California over the 1990-1999 period, and estimates of these rates in the US as a whole. Additionally, we have adjusted these rates so that they are equal to the California rates in 1989. These rates provide factual and counterfactual smoking behavioral relationships in California over the 1990-1999 period in the presence of and in the absence of California's Tobacco Control Program (TCP). However, in fact, the consequences of smoking, in terms of medical resources attributable to smoking, cases of smoking attributable diseases, health status, person years of life saved, and the value of years of life saved, take a life time to be revealed. So, to estimate the value generated by California's TCP we need to estimate the full life consequences to California residents operating with smoking behaviors obtained with and without California's TCP.

We estimated these results by simulating the life-time outcomes for California's 1990 residents twice, once using the observed smoking initiation and quitting rates and once using the adjusted national smoking initiation and quitting rates. We then compared the outcomes from the two simulations and attributed the estimated difference to California's Tobacco Control Program. The basic information for this evaluation is: (1) a description of the California population, a sample, describing the age and smoking behavior of its residents in the base year 1990; (2) age specific estimates of the smoking initiation and quit rates over the decade of the nineties; (3) models to estimate probabilities of relevant events--death, disease, health status, and costs (given disease status) in the simulations, and random processes to convert the probabilities into events determining the calculation of the comparative simulated outcomes. We seek to estimate the distribution of the population's simulated outcome arising from the different smoking conditions and the random processes. However, owing to the large standard errors when the TCP evaluation is based directly on the California 1990 tobacco sample, we adopted a cell-replication design to represent that sample. The purpose of this appendix is to explain how we went about obtaining results from this comparative simulation design.

The following notation will help provide structure to the argument: X_s denotes a vector of the relevant characteristics of individuals in the sample describing the California population in the base year of the simulations; Y[s] denotes an outcome over the simulation based on sample observations; w[s] denotes the number of people in the population represented by sample observation s; and Ns is the number of observations in the sample. If the sample were used as the basis for comparing the simulations, the expected value of the outcome Y[s] is estimated by Ybar[s]. The estimate of the variance of Ybar[s] is estimated by $(S^2)_{\text{Ybar}}[s]$. We can calculate Y[s], but how do we estimate Ybar[s]? We can calculate $S^2$for the population from the sample, but how do we estimate $S^2{}_{\text{Ybar}}[s]$? In more detail, then, a principal purpose of this appendix is to address these questions efficiently and by addressing questions efficiently I mean for the calculation effort expended--that is, how do we create minimum variance estimates of the distribution of outcome in the population.

- **A Computer Experiment**

Our strategy is to reformulate the information in the sample into a designed computer experiment and to estimate the answers sought from the experiment. Then, to transfer the knowledge back to the sample, which is then used to estimate the distribution of outcome in the population (Santner, Williams, Notz, 2003). To accomplish this, we construct a design representation of the population as described by the sample. We partition the space describing the relevant population characteristics into Nc disjoint cells. Then we represent the individuals in each cell with a prototypical individual with characteristics X_i, i=1,...Nc. Since any sample member will be subjected to the random processes required by simulations, understanding the distribution of outcome resulting from these processes requires replications of each prototypical individual in each cell i. Let J[i] denote the number of replications in cell i. The term w[i] is the number of people in the population represented by cell i. It is estimated by counting up the weights attached to the sample members who would occupy cell i. $S_{\text{Ybar}}[X[i]]$ is the standard deviation of the average outcome score for cell i. It is estimated by

$$S_{\text{Ybar}}[X[i]] = \sum_{i=1}^{Nc} (Y[X[i]] - \text{Ybar}[X[i]])^2 \Big/ (J[i]\,(J[i]-1)).$$

Make J[i] simulations, calculate Ybar[X[i]] and $(S^2)_{\text{Ybar}}[X[i]]$. The product of w[i] and $(S^2)_{\text{Ybar}}[X[i]]$ estimates the variance of the outcome score in the population derived from cell i.

The term $\text{Sqrt}\left[\sum_{i=1}^{Nc} w[i]\,(S^2)_{\text{Ybar}}[X[i]]\right]$ is the estimate of the total standard deviation of the outcome score in the population represented by the designed computer experiment. If C denotes the total number of simulations to be made in the experiment, one for each replication in the design, how many replications should be made in each cell so as to minimize the estimate of the variance in the distribution of outcome in the population?

Theorem: On the efficient allocation of replications in a computer experiment.

The efficient allocation of C replications across the Nc cells follows from choosing the number of replications for each cell, J[i], according to cell i's fraction of the total standard deviation of the outcome score in the population. That is,

$$J[i] = C\,\sqrt{w[i]}\,\hat{S}[i] \Big/ \sum_{i=1}^{Nc} \sqrt{w[i]}\,\hat{S}[i].$$

Proof:

Since each cell i is represented by a single replication's description X_i , we can suppress the dependence of outcome on characteristics. The estimate of the average and variance of the population outcome from cell i is given by equations [1] and [2]:

[1]

$$\text{Ybar}_{\text{in Population}} = \sum_{i=1}^{Nc} (w[i]\,\text{Ybar}[i]) \Big/ \sum_{i=1}^{Nc} w[i]$$

[2]

$$(S^2)_{\text{in Population}} = \sum_{i=1}^{Nc} w[i]\,(S^2)_{\text{Ybar}}[i]$$

where:

Ybar[i] = $\sum_{j=1}^{J[i]} Y[i, j] / J[i];$

$S^2[i] = \sum_{j=1}^{J[i]} \frac{(Y[i,j]-Ybar[i])^2}{-1+J[i]}$.; and

$(S^2)_{Ybar}[i] = S^2[i]$ /J[i]

To simplify the exposition, assume $S^2[i]$ is estimated independent of the J[i] determination process. For example, a two step estimation of $(S^2)_{Ybar}[i]$ is made. First, with a relatively small sample, $S^2[i]$ is estimated for the purpose of understanding how replications should be allocated to cells, and then $(S^2)_{Ybar}[i]$ is estimated with the J[i] replications for the purpose of furthering the outcome analysis.

The Lagrangian of the variance minimization expresses the objective function, the variance arising in the Nc cells, subject to the conditions that the sum of the replications in all the cells equals the number of calculations C and that the sum of the weights in the cells equals the population size. The Lagrangian is as follows:

$$\mathfrak{L} = \sum_{i=1}^{Nc} \frac{w[i]\, S^2[i]}{J[i]} + \lambda \left(-C + \sum_{i=1}^{Nc} J[i]\right) + \mu \left(-P + \sum_{i=1}^{Nc} w[i]\right)$$

The optimization problem is to minimize $\mathfrak{L}$ over the choice of the set J[i].

Taking the partial derivative of the Lagrangian with respect to J[i], and setting it to zero yields the first of the first order conditions for the estimate of the variance of outcome,

$$[3] \quad \lambda - \frac{w[i]\, S[i]^2}{J[i]^2} = 0$$

Taking the partial derivative of the Lagrangian with respect to the first constraint, λ, and setting it to zero yields the second of the first order conditions for the estimate of the variance of outcome,

[4] $-C + \sum_{i=1}^{Nc} J[i] = 0$

and taking the partial derivative of the Lagrangian with respect to the second constraint, μ, and setting it to zero yields the third of the first order conditions for the estimate of the variance of outcome,

[5] $-P + \sum_{i=1}^{Nc} w[i] = 0$.

Equation [3] actually represents Nc first order conditions of the form

[6] $\lambda = \frac{S^2[i]\, w[i]}{J[i]^2}$,

which all have the following solution for the optimal number of replications, J[i],

$$[7]\quad \text{Solve}\left[J[i]^2 - (1/\lambda)\ S^2[i]\ w[i] == 0,\ J[i]\right]$$

$$\left\{\left\{J[i] \to -\frac{\sqrt{S^2[i]}\ \sqrt{w[i]}}{\sqrt{\lambda}}\right\},\ \left\{J[i] \to \frac{\sqrt{S^2[i]}\ \sqrt{w[i]}}{\sqrt{\lambda}}\right\}\right\}$$

Accepting the positive valued solution with a positive square root yields equation [8],

[8] $J[i] = \frac{\sqrt{w[i]}\ \ S[i]}{\sqrt{\lambda}}$.

Now incorporate the second of the first order conditions, equation [4], $\sum_{i=1}^{Nc} J[i] = C$, into the analysis. Substituting the solution for J[i] into equation [4], obtains equation [9],

$$[9]\quad -C + \sum_{i=1}^{Nc} \frac{\sqrt{w[i]}\ \ S[i]}{\sqrt{\lambda}} == 0$$

which can be solved for $\sqrt{\lambda}$,

[10]

$$\sqrt{\lambda} = \sum_{i=1}^{Nc} \frac{\sqrt{w[i]}\ \ S[i]}{C}$$

Substitute this solution for $\sqrt{\lambda}$ back into the solution for J[i] (equation [8]) and we have proved our theorem,

[11] $J[i] = C\left(\sqrt{w[i]}\ S[i] \Big/ \sum_{i=1}^{Nc} \sqrt{w[i]}\ S[i]\right)$.

- **An algorithm to determine the optimal number of representations in a cell.**

The estimation of J[i] requires estimates of w[i] and $\hat{S}$[i]. By adding up the weights of every one in the sample represented by cell i we estimate w[i]. That is,

[12] w[i] = $\sum_{j=1}^{J[i]} w[s|s \epsilon i]$.

Take a reasonable, but small number of replications for every cell. Perhaps 30. Conduct the simulations for each cell and from the resulting outcome measures, estimate S[i] with the use of equations [1] and [2].

Based on these estimates for w[i] and S[i], for every i, compute $\sqrt{w[i]}\ S[i]$ and $\sum_{i=1}^{Nc} \sqrt{w[i]}\ S[i]$, and then the fraction of the contribution of cell i to the standard deviation in the total outcome, fJ[i], is given by equation [13].

[13] fJ[i] = $\frac{\sqrt{w[i]}\ S[i]}{\sum_{i=1}^{Nc} \sqrt{w[i]}\ S[i]}$.

Having decided to make C calculations and hence requiring C replications, the number for cell i is simply the product of fJ[i] and C,

[14] J[i] = fJ[i] C.

- **How many calculations, C, should be made?**

Let us assume that at the end of the analysis we desire a coefficient of variation (σ/μ) to have an estimated value (S/Ybar) equal to α. The coefficient of variation, estimated by $\left(\left(S^2\right)_{\text{in Population}}\right)^{1/2} / \mathbf{Ybar}_{\text{in Population}}$, where these terms are given by equations [2] and [1], respectively.

From estimates of equation [2],

[15] $\left(\hat{S}^2\right)_{\text{in Population}} = \sum_{i=1}^{Nc} \left(w[i]\ S^2[i] \big/ J[i]\right)$,

substitute in the value of J[i] from equation [11]]. The variance in the population is given by equation [16],

[16] $\left(S^2\right)_{\text{in Population}} = \sum_{i=1}^{Nc} \left(\left(w[i]\ S^2[i]\right) \Big/ \frac{\sqrt{w[i]}\ S[i]}{\sum_{i=1}^{Nc} \sqrt{w[i]}\ S[i]}\ C\right)$.

Simplify, and then solve for C.

$$C = \left(\sum_{i=1}^{Nc} \left(\sqrt{w[i]}\ S[i]\right)\right)^2 \Big/ \left(S^2\right)_{\text{in Population}}$$

The coefficient of variation, denoted by α, is estimated by $(\hat{S})_{\text{in Population}}$ /Ybar $_{\text{in population}}$

the [17] $C = \left(\sum_{i=1}^{Nc} \left(\sqrt{w[i]}\ S[i]\right)^2 \Big/ \left(\alpha^2\ \mathrm{Ybar}_{\text{in Population}}{}^2\right)\right.$

An algorithm to determine the required number of calculations.

The w[i] values are data and the initial estimates of $\hat{S}$[i] and Ybar[i] are obtained from the initial experiment. Employing equation [1] yields an estimate of $\mathbf{Ybar}_{\mathbf{in\ Population}}$ (= $\sum_{j=1}^{J[i]}$[w[i] Ybar[i]). For a given value of α, C is calculated from equation [17] and distributed among the Nc cells according to equation [14].

- **An analysis of the designed computer experiment. A transformation of the experiment information into sample knowledge.**

At this point assume the computer experiment has been conducted and we have obtained a vector of average outcomes for the cells, Ybar[i], and a vector of the standard deviations in outcome for the cells, $S_{\text{Ybar}}[i]\}$. The task now is to transform these statistics about the computer experiment into knowledge about the sample that can be used to estimate knowledge about the population.

We relate the statistics from the computer experiment with a multiplicative heteroscedastic regression model. This is the specification examined in depth by Harvey (1976), but our formulation is different because Harvey had no estimates of the variance of Ybar[i] and we do. And, accordingly, our results will differ from his. The model has the form specified by equation [18],

[18] $\text{Ybar[i]} = \text{X[i]}\,\beta + \epsilon[i]$

$$
\begin{aligned}
&\text{E}[\epsilon\text{[i]}\,\epsilon\text{[i]'}] = \sigma^2[i] = \\
&\quad = \text{Exp[Z[i]}\gamma\text{]=Exp[}\gamma\text{0] Exp[Z1[i] }\gamma\text{1] ...Exp[Zp[i]}\gamma\text{p]} \\
&\quad = \sigma_0{}^2[i]\ \text{Exp[Z1[i] }\gamma\text{1] ...Exp[Zp[i]}\gamma\text{p]} \\
&\quad = \sigma_0{}^2[i]\ \text{Exp[Z*[i] }\gamma\text{*]}
\end{aligned}
$$

for all i, where:

X[i] is a row vector, 1 x P, of the P descriptive characteristics of the prototypical member of cell i,

β is a vector of length P,

γ is a vector of length K, and

ϵ[i] is a random variable indicating the difference between cell [i]'s average outcome and the cell i's expected outcome, given its characteristics, X[i].

If Mϵ is a vector of the cell's random variables, MYbar is a vector of the cell average outcomes, MX is a Nc by P matrix of the cell characteristics, relevant to describing Ybar and MZ is a Nc by K matrix of the cell characteristics relevant to describing the variance in Ybar, then the expected value vector, variance-covariance matrix, and estimate of the variance-covariance matrix is given by equations [19a], [19b], and [19c] respectively,

[19a] $\text{E[M}\epsilon\text{]=0}$; and

[19b] $\text{E[M}\epsilon\text{M}\epsilon\text{']=}$

$$
\begin{array}{cccc}
\text{Exp}[Z[i=1]\gamma], & 0\,, & \ldots\,, & 0 \\
0\,, & \text{Exp}[Z[i=2]\gamma], & \ldots\,, & 0
\end{array}
$$

$$\Sigma = [\begin{array}{cccc} 0, & 0, & \ldots, & \mathrm{Exp}[Z[i = \mathrm{Nc}]\,\gamma] \end{array}].$$

[19c] $M(S^2)_{\mathrm{Ybar}} = \Psi =$

$$= [\begin{array}{cccc} (S^2)_{\mathrm{Ybar}}[X[i=1]], & 0, & \ldots, & 0 \\ 0, & (S^2)_{\mathrm{Ybar}}[X[i=2]], & \ldots, & 0 \\ 0, & 0, & \ldots, & (S^2)_{\mathrm{Ybar}}[X[i=\mathrm{Nc}]] \end{array}].$$

Since estimates of the variance are known, generalized least squares provides minimum variance estimates b of β.

[20] b = (X[i]' Ψ^{-1}X[i]) (X[i]' Ψ^{-1}Y[i]).

We turn now to the estimation of the γ coefficients. Based on the estimates b of β, an observed error in the model for Ybar[i] is given by

[21] e[i] = Ybar[i] - X[i] b.

The logarithm of the square of this observed error is the estimate of the variance for an observation, which by the postulated multiplicative heteroscedastic model, is

[22] Log[$e[i]^2$] = Z[i] γ + v[i].

In our case, we have an estimate of this variance, so equation [21] can be expressed as equation [22],

[23] Log[$(S^2)_{\mathrm{Ybar}}$[i]] = Z[i] γ + v[i].

Let c denote the least squares estimator of γ. c is given by equation [24],

[24] c = $(Z'\,Z)^{-1}$Z' Log[$(S^2)_{\mathrm{Ybar}}$[i]].

We now examine the characteristics of this estimator. Our analysis is similar to that of Harvey (1976), though somewhat simpler because the dependent variable, Log[$(S^2)_{\mathrm{Ybar}}$[i]], is observed. From equation [17], Z[i] γ = Log $\sigma^2[i]$. After substituting into equation [23], and solving for the error term, we have equation [25],

[25] ν[i] = Log[$\left(S^2\right)_{\text{Ybar}}$[i]] - ln $\sigma^2[i]$= Log [$\left(S^2\right)_{\text{Ybar}}$[i]] / $\sigma^2[i]$]

Under the assumption that the deviations from the means of a cell are Normal, Log [$\left(S^2\right)_{\text{Ybar}}$[i]] / $\sigma^2[i]$] is distributed as the natural logarithm of a Chi-Squared distribution with J[i] degrees of freedom divided by J[i] (recall J[i] are the number of observations in cell [i]) the error term v[i] is so distributed.

The expected value of the Logarithm of a Chi-Squared with one degree of freedom equals -1.27036.

```
Integrate[Log[x] PDF[ChiSquareDistribution[1], x], {x, 0, Infinity}]

-EulerGamma - Log[2]

N[%]

-1.27036
```

By subtracting the expected value of the error from the constant, c0, and from the error v[i], we obtain,

$$[26]\quad E[c] = \gamma + (Z'Z)^{-1} Z' E[\text{Log}[v[i]]] \\ = \gamma + (Z'Z)^{-1} Z' (-1.27036)$$

which implies

$$[27]\quad E[\tilde{c}] = [c - 1.27036\, i] = \\ = \gamma + (Z'Z)^{-1} Z' E[\text{Log}[v[i] - 1.27036]] \\ = \gamma$$

where i = {1,0,...,0}', of length P.

The variance of a random variable distributed as the logarithm of a Chi-Squared with one degree of freedom has a value equal to $\frac{\pi^2}{2}$, which equals 4.9348.

```
[28] Integrate[
  (Log[x] - (-EulerGamma - Log[2]))^2 PDF[ChiSquareDistribution[1], x], {x, 0, Infinity}]
```

$$\frac{\pi^2}{2}$$

```
N[%]

4.9348
```

The variance of $\tilde{c}$ is given by equation [29],

$$[29]\quad V[c] = V\left[(Z'Z)^{-1} Z'(Z\gamma + v[i])\right]$$

$$= (Z'Z)^{-1} (Z'\ V[v[i]]\text{Z})\ (Z'Z)^{-1}$$

$$= (Z'Z)^{-1}\ 4.9348$$

- **Applying the cell analysis to the sample**

We turn now to use the estimated models to estimate the mean outcome and its variance, given the sample data.. Let X[s] and Z[s] describe the relevant model characteristics with a sample member s. We wish to forecast the distribution of the average outcome for sample member s and its variance. The average outcome forecast for sample member s is given by equation [30],

[30] $\hat{Y}\,\mathrm{bar}[s] = X[s]\,b + \epsilon[s]$,

where ϵ[s] is the forecast error. The Gauss-Markov theorem insures that the minimum variance linear unbiased estimator of the forecasted average outcome for sample member s , $\hat{Y}$bar[s], is given by equation [31],

[31] $\hat{\hat{Y}}\,\mathrm{bar}[s] = X[s]\,b$.

Let $S^2{}_{\text{forcast}}$ denote an estimate of the variance in the forecast error. Based on the cell data, we estimate the variance of the forecast error with equation [32],

[32] $\hat{S}^2{}_{\text{forcast error}} = \sum_{i=1}^{Nc} \left(-b\,X[i] + \hat{Y}\,\mathrm{bar}[i]\right)^2$/(Nc-1).

The estimate of the variance of the mean of a sample observation, given X[s], is based on the models estimated and the sample information. Let $(\hat{\hat{S}}^2)_{\text{Ybar}}$[s] represent the estimate of the variance of a forcasted value of Ybar based on sample data. This variance should be the sum of the estimated variance of Ybar, given sample data, plus the variance of the forecast error. The expression for $(\hat{\hat{S}}^2)_{\text{Ybar}}$[s], simply combines the predicted value of the estimate of the variance of Ybar and the average forecast error calculated with equation [32],

[33] E[$(\hat{\hat{S}}^2)_{\text{Ybar}}$[s]]= E[Exp[Z[s] γ + v[s]] + $\hat{S}^2{}_{\text{forcast error}}$

= E[Exp[Z[s] $\tilde{c}$]] + $\hat{S}^2{}_{\text{forcast error}}$ =

An example of a model specification

To illustrate the ideas developed above, we offer an example of a specification for the model presented by equation [18]. The specification of X[i] might be described by equation [34] :

[34] $X[i] = \{1, X1[i], X1[i]^2, X1[i]^3, X22[i], X23[i], X3[i], X4[i], X5[i]\};$

$\beta = \{\beta 0, \beta 11, \beta 12, \beta 13, \beta 22, \beta 23, \beta 3, \beta 4, \beta 5\}';$

and the individual X elements are defined as follows:
X1=age;
X2={1/0} according as observation is an {ever-smoker in 1990/otherwise};
X22={1/0} according as observation is a {former smoker in 1990/otherwise};
X23={1/0} according as observation is a {current smoker in 1990/otherwise};
X3=start age of a current or former smoker in 1990;
X4=quit age of a former smoker in 1990;
X5=cigarettes smoked per day,Modulo 1/2.

and the specification of the Z might be:

[35] $Z[i]= \{1, X1[i], X1[i], X2[i], X4[i], X5[i]\}$

and accordingly, $\gamma = \{\gamma 0, \gamma 11, \gamma 12, \gamma 2, \gamma 4, \gamma 5\}'$.

The Evaluation of the Program with sample knowledge.

Having estimated the average outcome and its variance for every sample member, the TCP program is then evaluated. The outcome in the population is estimated from the sample data with equation [36],

[36] $Y_{\text{in Population}} = \sum_{s=1}^{Ns} \hat{\hat{Y}} \text{bar}[s]\, w[s],$

and the variance of the outcome in the population is estimated from the sample data with equation [37],

[37] $\left(S^2\right)_{\text{in Population}} = \sum_{s=1}^{Ns} w[s]^2 \left(\hat{\hat{S}}^2\right)_{\text{Ybar}}[s].$

In fact, since we have many different outcomes of interest, a particular one must be choosen to determine the design structure. We choose years of life saved and the design criteria.

- **Optimum allocation of simulation calculations.**

1. Allocate 30 replications to each cell, generate the simulation outcomes, and estimate the average and variance of these cell outcomes, Ybar[i] and $S^2[i]$, for each cell.

2. Compute w[i] for each cell.

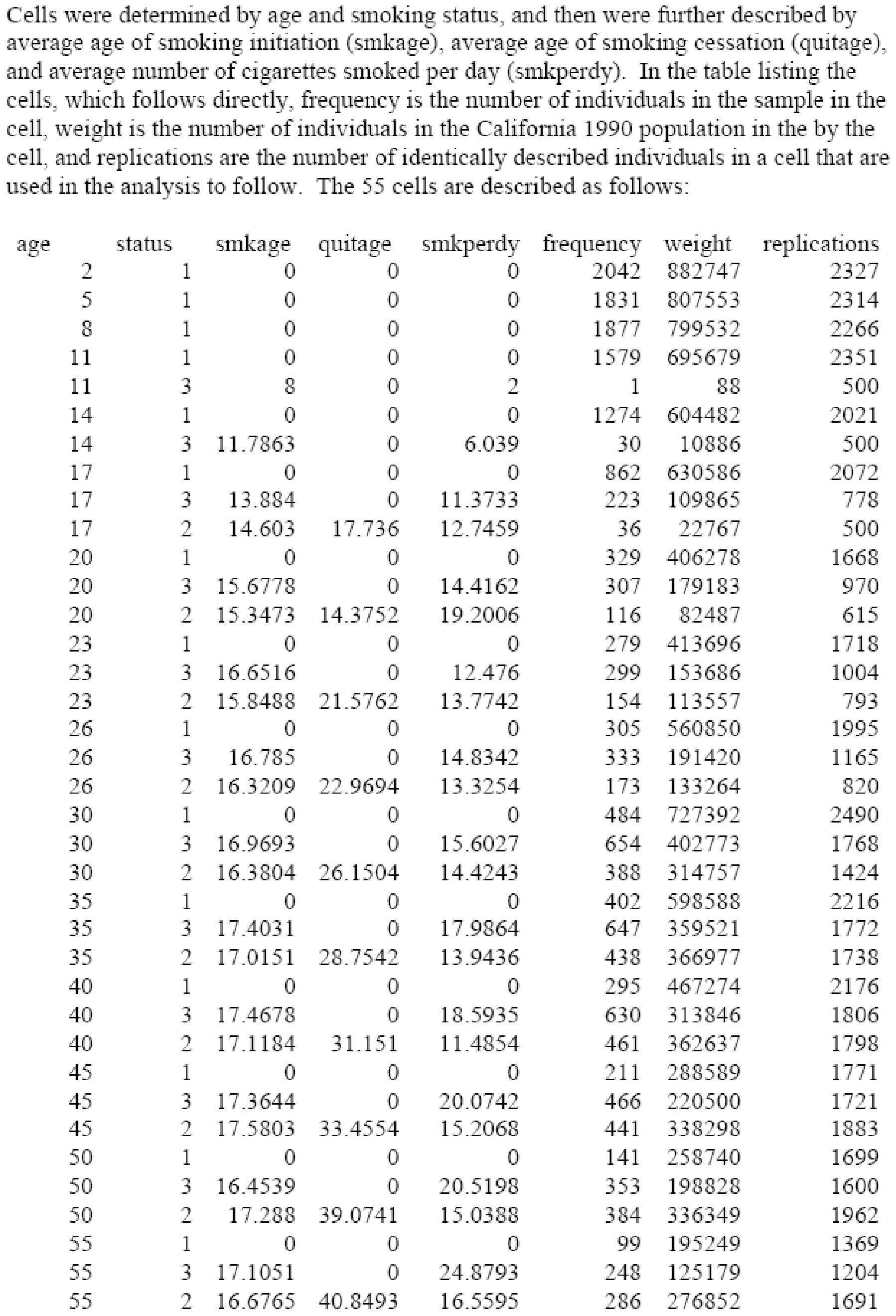
3. Estimate the set {Ji} of replications per cell.

The following algorithm is then implemented to obtain a simulation result.

- **1. Model Ybar[i]**
- **2. Model $(S)_{\text{Ybar}}[i]$**
- **3. Estimate the average outcomes in the sample, based on cell data.**
- **4. Estimate the variance of the average outcomes in the sample, based on cell data.**
- **5. Estimate the average outcomes in the population, based on sample data.**
- **6. Estimate the variance of the average outcomes in the population, based on sample data.**
- **7. Report the point and interval estimates of the outcomes attributable to TCP.**

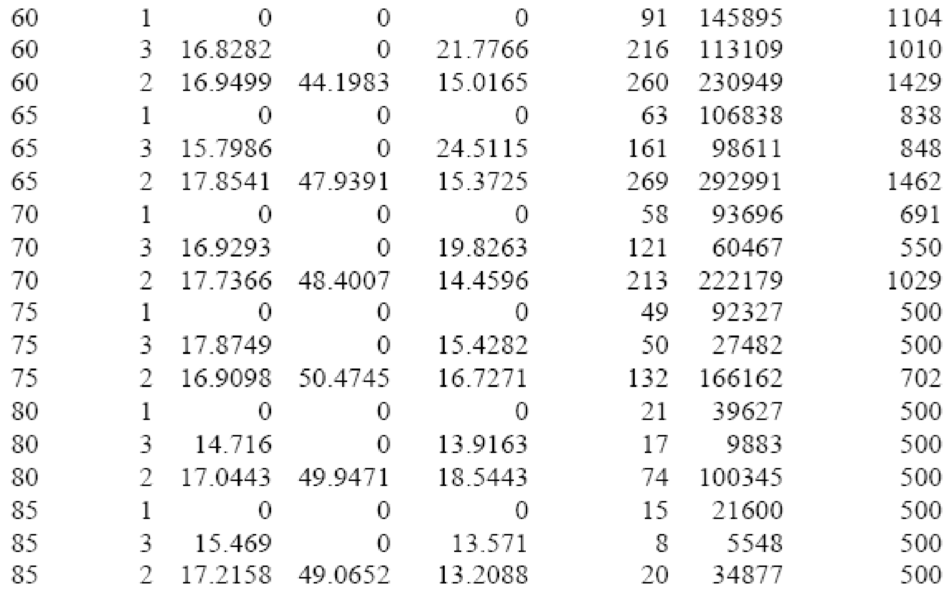
- **Implementation of the cell-replication design.**

In the study at hand, a cell is described with five values in 1990:
X1=age, range 1 to 90 ;
X2=smoking status 1,2,3;
X3=start age 11 to 22 ;
X4=quit age 20 to 90 ; and
X5=smk packs/day 0 to5/2, in units of 1/2 packs per day, 0-1/2, 1/2-1,...

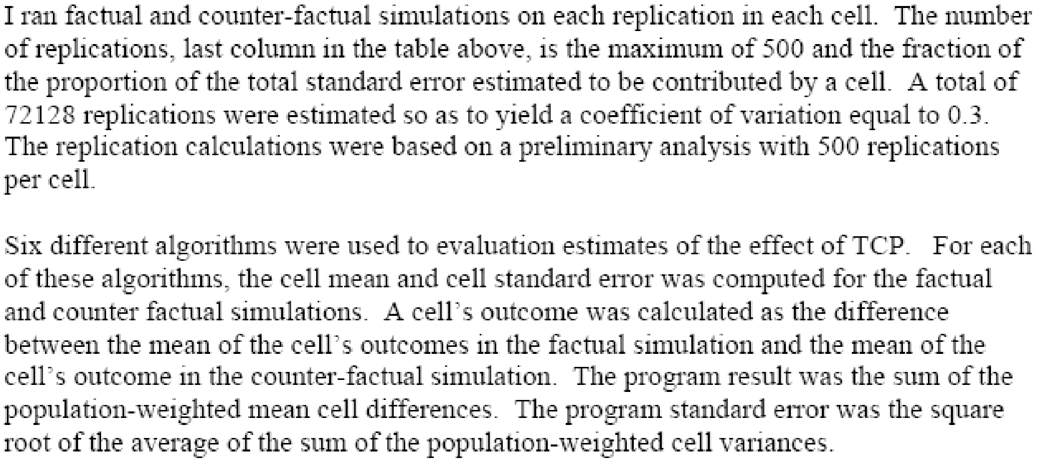


Cells were determined by age and smoking status, and then were further described by average age of smoking initiation (smkage), average age of smoking cessation (quitage), and average number of cigarettes smoked per day (smkperdy). In the table listing the cells, which follows directly, frequency is the number of individuals in the sample in the cell, weight is the number of individuals in the California 1990 population in the by the cell, and replications are the number of identically described individuals in a cell that are used in the analysis to follow. The 55 cells are described as follows:

age	status	smkage	quitage	smkperdy	frequency	weight	replications
2	1	0	0	0	2042	882747	2327
5	1	0	0	0	1831	807553	2314
8	1	0	0	0	1877	799532	2266
11	1	0	0	0	1579	695679	2351
11	3	8	0	2	1	88	500
14	1	0	0	0	1274	604482	2021
14	3	11.7863	0	6.039	30	10886	500
17	1	0	0	0	862	630586	2072
17	3	13.884	0	11.3733	223	109865	778
17	2	14.603	17.736	12.7459	36	22767	500
20	1	0	0	0	329	406278	1668
20	3	15.6778	0	14.4162	307	179183	970
20	2	15.3473	14.3752	19.2006	116	82487	615
23	1	0	0	0	279	413696	1718
23	3	16.6516	0	12.476	299	153686	1004
23	2	15.8488	21.5762	13.7742	154	113557	793
26	1	0	0	0	305	560850	1995
26	3	16.785	0	14.8342	333	191420	1165
26	2	16.3209	22.9694	13.3254	173	133264	820
30	1	0	0	0	484	727392	2490
30	3	16.9693	0	15.6027	654	402773	1768
30	2	16.3804	26.1504	14.4243	388	314757	1424
35	1	0	0	0	402	598588	2216
35	3	17.4031	0	17.9864	647	359521	1772
35	2	17.0151	28.7542	13.9436	438	366977	1738
40	1	0	0	0	295	467274	2176
40	3	17.4678	0	18.5935	630	313846	1806
40	2	17.1184	31.151	11.4854	461	362637	1798
45	1	0	0	0	211	288589	1771
45	3	17.3644	0	20.0742	466	220500	1721
45	2	17.5803	33.4554	15.2068	441	338298	1883
50	1	0	0	0	141	258740	1699
50	3	16.4539	0	20.5198	353	198828	1600
50	2	17.288	39.0741	15.0388	384	336349	1962
55	1	0	0	0	99	195249	1369
55	3	17.1051	0	24.8793	248	125179	1204
55	2	16.6765	40.8493	16.5595	286	276852	1691



60	1	0	0	0	91	145895	1104
60	3	16.8282	0	21.7766	216	113109	1010
60	2	16.9499	44.1983	15.0165	260	230949	1429
65	1	0	0	0	63	106838	838
65	3	15.7986	0	24.5115	161	98611	848
65	2	17.8541	47.9391	15.3725	269	292991	1462
70	1	0	0	0	58	93696	691
70	3	16.9293	0	19.8263	121	60467	550
70	2	17.7366	48.4007	14.4596	213	222179	1029
75	1	0	0	0	49	92327	500
75	3	17.8749	0	15.4282	50	27482	500
75	2	16.9098	50.4745	16.7271	132	166162	702
80	1	0	0	0	21	39627	500
80	3	14.716	0	13.9163	17	9883	500
80	2	17.0443	49.9471	18.5443	74	100345	500
85	1	0	0	0	15	21600	500
85	3	15.469	0	13.571	8	5548	500
85	2	17.2158	49.0652	13.2088	20	34877	500



I ran factual and counter-factual simulations on each replication in each cell. The number of replications, last column in the table above, is the maximum of 500 and the fraction of the proportion of the total standard error estimated to be contributed by a cell. A total of 72128 replications were estimated so as to yield a coefficient of variation equal to 0.3. The replication calculations were based on a preliminary analysis with 500 replications per cell.

Six different algorithms were used to evaluation estimates of the effect of TCP. For each of these algorithms, the cell mean and cell standard error was computed for the factual and counter factual simulations. A cell's outcome was calculated as the difference between the mean of the cell's outcomes in the factual simulation and the mean of the cell's outcome in the counter-factual simulation. The program result was the sum of the population-weighted mean cell differences. The program standard error was the square root of the average of the sum of the population-weighted cell variances.

Bibliography

**Santner, TJ, B.J.Williams, and W.I.Notz,
2003. The Design and Analysis of Computer Experiments.Springer.New York.**