**TRANSACTIONS OF SOCIETY OF ACTUARIES**
**1992 VOL. 44**

# MODELING FLEXIBLE BENEFIT SELECTION

CHARLES S. FUHRER AND ARNOLD F. SHAPIRO

ABSTRACT

A mathematical framework for benefits and choices must be created to model benefit selection. This paper creates such a framework by defining benefit plans as reimbursement functions. These are then used with a defined choice function to calculate the cost deviation due to selection. Finally, utility functions can be applied to this framework to predict choice.

## I. INTRODUCTION

The problem of selection has been recognized by actuaries since the early days of the profession and has been a continuing concern since then. Highan [14] in 1851, for example, authored an article in the first volume of the *Journal of the Institute*, entitled "On the Value of Selection as Exercised by the Policy-holder against the Company." Similarly, McClintock [19] in 1892, in an early volume of the *Transactions of the Actuarial Society of America*, published an actuarial essay "On the Effect of Selection."

During the early periods, the analysis was primarily descriptive and concerned with identifying situations conducive to adverse selection and the associated hazards. In recent years, the emphasis has changed towards an attempt to model the selection process and an analysis of the sensitivity of those models. Moreover, while the initial concern was raised by actuaries in the context of insurance, it has come to be recognized as an issue common to a number of commodities and, as such, has become an important field of study in economics.

A number of issues have emerged. The optimal form of an insurance contract for a risk-averse insured was studied by Borch [5], Arrow [2], Raviv [22], Bühlmann and Jewell [7], and Blazenko [4]. Models that addressed the difficulty created by asymmetric market information on the riskiness of the insured were developed by Akerlof [1], Rothchild and Stiglitz [23], Wilson [25], Miyazaki [20], and Spence [24]. Still others have studied the role of wealth in this decision process: Gould [13], who concluded that it was not appropriate to consider demand without regard for the wealth position of the individual; Mayers and Smith [18]; and Doherty and Schlesinger [11], who showed how assets are correlated with the demand for insurance.

This paper extends the analysis by dealing with some of the statistical aspects of choice in benefit plans. Although these techniques could be used for any choice in insurance plans, the focus is on group health benefit plans. By group health benefit plan we mean a system in which the members of a group are eligible to receive insurance benefits for some part of the cost of their (and sometimes their family's) medical care. The insurance benefits may require the payment of premiums. In general, the particular plan of benefits and premiums is unique to each group. The group is usually formed for some purpose other than the insurance coverage. The most common groups are the employees of a single employer.

Most of the remarks deal with the traditional health insurance indemnity plans in which group members obtain health care from licensed health care providers and then are reimbursed for a portion of the providers' charges. Some benefit plans include a provision for an employee choice between more than one formula for the amount of reimbursement. The employee may be required to contribute different premiums for each option.

Employee choice in group health benefits has started to become popular only in the last 5 or 10 years in the U.S. Of course, trivially, most plans have always allowed the choice of rejecting the coverage if the employee is required to pay premiums for the coverage. Thus, there is a choice between the benefit plan and a null plan.

## II. REIMBURSEMENT

Before we can write some expressions for the effects of selection or predict it, we need to express the whole set of choices and outcomes in a functional and probabilistic setting.

Let the random variable $X$ be the covered charges for an individual during a period, usually one year. Assume that $X$ is a one-dimensional positive random variable.

We define the notation:

$$x^+ = \max\{0,x\} = \begin{cases} 0 & x < 0 \\ x & x \geq 0 . \end{cases}$$

Let $r(X)$ be the amount of reimbursement in a benefit plan for covered charges equal to $X$, where $r$ is a function called here a reimbursement function. Note that we are assuming that the amount of reimbursement is determined only by the total of covered charges during the year and not by when the services were performed or by which providers.

Although any function $r$ could be a reimbursement function, we note that, in general, they have the following properties:

I. They are continuous: $\lim_{x \to a} r(x) = r(a)$;

II. They are nondecreasing: $x > y \Rightarrow r(x) \geq r(y)$;

III. $x > y \Rightarrow r(x) - r(y) \leq x - y$; and

IV. $r(0) = 0$.

Property I says that the amount reimbursed cannot vary too much for small changes in covered charges. Property II says that as the covered charges increase, the reimbursement cannot decrease. Property III says that amount of reimbursement cannot increase faster than covered charges. Property IV says that there is no reimbursement when there are no covered charges.

*Example 2.1*

The reimbursement function can be the identity function: $r(x) = x$. This is full reimbursement for all covered charges.

*Example 2.2*

The reimbursement function can be identically equal to zero: $r(x) = 0$ for all $x$. This is the case of no benefits.

*Example 2.3*

For a given fixed constant $d$,

$$r(x) = (x - d)^{+} = \begin{cases} 0 & x \leq d \\ x - d & x > d. \end{cases}$$

This is called full coverage after a deductible. The constant is called the deductible.

*Example 2.4*

For a constant $c$, $0 < c < 1$, $r(x) = cx$. The constant is called the coinsurance rate.

*Example 2.5*

We can have both a deductible and coinsurance (a combination of examples 2.3 and 2.4):

$$r(x) = c(x - d)^+ = \begin{cases} 0 & x \leq d \\ c(x - d) & x > d. \end{cases}$$

*Example 2.6*

There can be a limit on the coinsurance of example 2.4. For constant $L > 0$ and $c$, $0 < c < 1$[1]:

$$r(x) = cx + [(1 - c)x - L]^+ = \begin{cases} cx & x < L/(1 - c) \\ x - L & x \geq L/(1 - c). \end{cases}$$

Here $L$ is known as the coinsurance limit or maximum. Note that $L$ is not the amount of covered charges that has to be reached before full reimbursement but rather is the maximum that is not reimbursed.

*Example 2.7*

Examples 2.5 and 2.6 can be combined to get a plan with deductible, coinsurance, and coinsurance limit:

$$r(x) = c(x - d)^+ + [(1 - c)(x - d) - L]^+$$

$$= \begin{cases} 0 & x < d \\ c(x - d) & d \leq x < L/(1 - c) + d \\ x - d - L & L/(1 - c) + d \leq x. \end{cases}$$

In this case $L + d$ is sometimes called the out-of-pocket limit.

*Example 2.8*

Often there is an overall individual annual benefit maximum. For a constant $M$,

[1]Note that we have deviated from the usual convention of reserving the uppercase for random variables.

$$r(x) = \min\{x, M\} = \begin{cases} x & x < M \\ M & x \geq M. \end{cases}$$

*Example 2.9*

A combination of examples 2.7 and 2.8 would be a plan with deductible, coinsurance, coinsurance maximum, and overall annual maximum:

$$r(x) = \min\{c(x - d)^+ + [(1 - c)(x - d) - L]^+, M\}$$

$$= \begin{cases} 0 & x < d \\ c(x - d) & d \leq x < L/(1 - c) + d \\ x - d - L & L/(1 - c) + d \leq x < M + d + L \\ M & M + d + L \leq x. \end{cases}$$

For this example, we define the intervals: $B = [d, L/(1-c)+d)$, $C=[L/(1-c)+d, M+d+L)$, and $D=[M+d+L,\infty)$. Even though this looks rather complicated, this is often just called a comprehensive major medical plan of benefits. Of course, examples 2.1 through 2.8 can be treated as special cases of this example 2.9. All the $r$'s in examples 2.1 through 2.9 satisfy the properties I through IV above.

Table 1 illustrates some sample $r$'s: $r_1$ is a very rich plan; $r_2$ reimburses less; $r_3$ is a cheap plan; $r_4$ is the null or 0 reimbursement of example 2.2; and $r_5$ is the full reimbursement of example 2.1.

TABLE 1

SOME SAMPLE REIMBURSEMENT FUNCTIONS

| Reimbursement Number ($r$) | Deductible ($d$) | Coinsurance ($c$) | Coinsurance Maximum ($L$) | Overall Annual Maximum ($M$) |
|---|---|---|---|---|
| 1 | \$ 100 | 80% | \$ 400 | \$1,000,000 |
| 2 | 500 | 80 | 1,000 | 1,000,000 |
| 3 | 1,000 | 75 | 3,000 | 500,000 |
| 4 | | 0 | | |
| 5 | 0 | 100 | | None |

*Example 2.10*

Assume that the random variable $X$ has the discrete distribution: $Pr\{X=ks\}=p_k$ for $k=0, 1, 2, \ldots$ and a constant $s$ called the unit or span.[2] Of course,

$$\sum_{k=0}^{\infty} p_k = 1.$$

Using the $r$'s of example 2.9, we can calculate some values:

$$E[r(X)] = \sum_{ks\in B} c(ks - d)p_k + \sum_{ks\in C} (ks - d - L)p_k + \sum_{ks\in D} Mp_k,$$

$$E[r^2(X)] = \sum_{ks\in B} c^2(ks - d)^2 p_k + \sum_{ks\in C} (ks - d - L)^2 p_k + \sum_{ks\in D} M^2 p_k,$$

and

$$\text{Var}[r(X)] = \text{E}[r^2(X)] - E^2[r(X)],$$

where we have used the notation: $r^2(X)=[r(X)]^2$ or $E^2(X)=[E(X)]^2$.

Table 2 shows an example of such a distribution. This distribution was based on data obtained from Health Care Service Corp. (Blue Cross/Blue Shield of Illinois).

Table 3 shows the expectation and variance of the five reimbursements of example 2.9 when using this distribution, with $s=\$1,000$.

*Example 2.11*

Similarly, let $X$ have the mixed distribution where $Pr\{X=0\}=p_0$ and $Pr\{a<x\leq b\}=\int_a^b f(t)dt$ for $a\geq 0$ and a density function $f$ such that $\int_0^\infty f(t)dt=1-p_0$. See Hogg and Klugman [15, page 50] for a discussion of mixed distributions. Again, assuming the $r$ of example 2.9, we have the values

$$E[r(X)] = \int_B c(t - d)f(t)\,dt + \int_C (t - d - L)f(t)\,dt + \int_D Mf(t)\,dt,$$

and

$$E[r^2(X)] = \int_B c^2(t - d)^2 f(t)dt + \int_C (t - d - L)^2 f(t)dt + \int_D M^2 f(t)dt.$$

[2]This formulation has the advantage of simplicity. An alternative formulation would be $Pr\{ks\leq X<(k+1)s\}=p_k$.

TABLE 2

SAMPLE DISCRETE DISTRIBUTION

| $s=1$: Mean = 1.433, Variance = 28.175, Standard Deviation = 5.308 | | | | | | | |
|---|---|---|---|---|---|---|---|
| $k$ | $p(k)$ | $k$ | $p(k)$ | $k$ | $p(k)$ | $k$ | $p(k)$ |
| 0 | 0.600839 | 43 | 0.000139 | 85 | 0.000023 | 131 | 0.000005 |
| 1 | 0.212998 | 44 | 0.000126 | 86 | 0.000019 | 132 | 0.000004 |
| 2 | 0.057230 | 45 | 0.000097 | 87 | 0.000023 | 133 | 0.000005 |
| 3 | 0.033316 | 46 | 0.000082 | 88 | 0.000015 | 134 | 0.000003 |
| 4 | 0.022218 | 47 | 0.000136 | 89 | 0.000005 | 135 | 0.000003 |
| 5 | 0.015504 | 48 | 0.000107 | 90 | 0.000011 | 136 | 0.000008 |
| 6 | 0.011159 | 49 | 0.000095 | 91 | 0.000017 | 137 | 0.000009 |
| 7 | 0.008179 | 50 | 0.000048 | 92 | 0.000018 | 138 | 0.000009 |
| 8 | 0.006329 | 51 | 0.000060 | 93 | 0.000009 | 139 | 0.000003 |
| 9 | 0.004906 | 52 | 0.000077 | 94 | 0.000004 | 140 | 0.000002 |
| 10 | 0.003751 | 53 | 0.000098 | 95 | 0.000006 | 142 | 0.000005 |
| 11 | 0.002734 | 54 | 0.000077 | 96 | 0.000015 | 145 | 0.000001 |
| 12 | 0.002257 | 55 | 0.000044 | 97 | 0.000007 | 146 | 0.000005 |
| 13 | 0.001984 | 56 | 0.000050 | 98 | 0.000021 | 147 | 0.000006 |
| 14 | 0.001629 | 57 | 0.000067 | 99 | 0.000014 | 148 | 0.000005 |
| 15 | 0.001230 | 58 | 0.000092 | 100 | 0.000005 | 150 | 0.000004 |
| 16 | 0.001179 | 59 | 0.000066 | 101 | 0.000013 | 151 | 0.000005 |
| 17 | 0.001041 | 60 | 0.000055 | 102 | 0.000015 | 152 | 0.000004 |
| 18 | 0.000854 | 61 | 0.000024 | 103 | 0.000015 | 153 | 0.000003 |
| 19 | 0.000741 | 62 | 0.000033 | 104 | 0.000012 | 158 | 0.000001 |
| 20 | 0.000633 | 63 | 0.000027 | 105 | 0.000011 | 159 | 0.000016 |
| 21 | 0.000554 | 64 | 0.000031 | 106 | 0.000003 | 160 | 0.000006 |
| 22 | 0.000529 | 65 | 0.000041 | 107 | 0.000004 | 169 | 0.000001 |
| 23 | 0.000528 | 66 | 0.000036 | 108 | 0.000007 | 170 | 0.000004 |
| 24 | 0.000485 | 67 | 0.000043 | 111 | 0.000002 | 172 | 0.000004 |
| 25 | 0.000397 | 68 | 0.000041 | 112 | 0.000007 | 173 | 0.000007 |
| 26 | 0.000387 | 69 | 0.000046 | 113 | 0.000005 | 185 | 0.000003 |
| 27 | 0.000352 | 70 | 0.000038 | 114 | 0.000007 | 186 | 0.000002 |
| 28 | 0.000403 | 71 | 0.000010 | 115 | 0.000006 | 197 | 0.000006 |
| 29 | 0.000333 | 72 | 0.000017 | 116 | 0.000001 | 202 | 0.000003 |
| 30 | 0.000306 | 73 | 0.000029 | 117 | 0.000009 | 203 | 0.000003 |
| 31 | 0.000253 | 74 | 0.000033 | 118 | 0.000002 | 204 | 0.000004 |
| 32 | 0.000258 | 75 | 0.000012 | 119 | 0.000005 | 205 | 0.000001 |
| 33 | 0.000245 | 76 | 0.000011 | 120 | 0.000004 | 206 | 0.000005 |
| 34 | 0.000228 | 77 | 0.000014 | 121 | 0.000005 | 245 | 0.000005 |
| 35 | 0.000204 | 78 | 0.000012 | 122 | 0.000010 | 263 | 0.000006 |
| 36 | 0.000231 | 79 | 0.000016 | 123 | 0.000004 | 285 | 0.000005 |
| 37 | 0.000193 | 80 | 0.000007 | 125 | 0.000002 | 292 | 0.000005 |
| 38 | 0.000172 | 81 | 0.000011 | 126 | 0.000003 | 323 | 0.000002 |
| 39 | 0.000177 | 82 | 0.000002 | 127 | 0.000005 | 324 | 0.000003 |
| 40 | 0.000133 | 83 | 0.000021 | 128 | 0.000013 | 519 | 0.000003 |
| 41 | 0.000121 | 84 | 0.000020 | 130 | 0.000005 | 520 | 0.000002 |
| 42 | 0.000136 | | | | | | |

TABLE 3

CALCULATION OF VALUES FOR THE REIMBURSEMENTS

| Reimbursement Number ($r$) | Discrete Distribution from Table 2; $s$ = $1,000 | | | Pareto Distribution | | |
|---|---|---|---|---|---|---|
| | Mean | Variance | Standard Deviation | Mean | Variance | Standard Deviation |
| 1 | $1,282.10 | 27,313,585 | $5,226.24 | $2,865.45 | 67,725,832 | $ 8,229.57 |
| 2 | 1,091.57 | 25,789,764 | 5,078.36 | 2,436.31 | 65,540,408 | 8,095.70 |
| 3 | 846.98 | 22,912,997 | 4,786.75 | 1,955.29 | 52,346,277 | 7,235.07 |
| 4 | 0.00 | 0 | 0.00 | 0.00 | 0 | 0.00 |
| 5 | 1,433.67 | 28,175,197 | 5,308.03 | 3,207.80 | 141,052,606 | 11,876.56 |

Table 3 also shows a calculation of these values using the Pareto distribution with the same mean and variance as the discrete distribution and $p_0 = 0$. The Pareto distribution is discussed in [9] and [15]; it is often used for claim size distributions. The Pareto has density: $f(x) = \alpha\lambda^{\alpha}(\lambda + x)^{-\alpha-1}$ and expectation of $\lambda/(\alpha - 1)$. The values used were: $\alpha = 2.1574$ and $\lambda = 1.6593$.

## III. COST DEVIATIONS DUE TO SELECTION

We assume that a group comprises $m$ individuals, $m \geq 1$. The covered charges for individual $i$ are denoted with the non-negative random variable $X_i$, $1 \leq i \leq m$. Now assume that each individual is given a choice at the beginning of the year between $n$ reimbursement functions: $r_1(x)$ ..., $r_n(x)$. To avoid long subscripts, we write $r_j(x) = r(j,x)$, $1 \leq j \leq n$. We define the "mean group reimbursement at $r_j$" as the random variable

$$\Psi(j) = \frac{1}{m} \sum_{i=1}^{m} r(j,X_i).$$

In the prechoice environment, insurers have been estimating $E[\Psi(j)]$ by using relatively complicated manual rating formulas that take into account the characteristics of the group, the individuals in the group, and $r_j$. The formulas are complicated because they must reflect the deductible, the coinsurance, and so on.[3] Incidentally, insurers often use the group's experience to estimate $E[\Psi(j)]$.

[3]Of course, this is not true for simple reimbursement functions such as in examples 2.1, 2.2, and 2.4, where $E[\Psi(j)] = r_j\left[\frac{1}{m} \sum_{i=1}^{m} E(X_i)\right]$.

Assume that the $i$-th member of the group, $1 \leq i \leq m$, chooses reimbursement level $\chi(i)$, $1 \leq \chi(i) \leq n$. Thus $\chi(i)$ is a function $\chi:\{1, 2, \ldots, m\} \rightarrow \{1, 2, \ldots, n\}$ called the choice function. Also, we define $P(j)$, $1 \leq j \leq n$ as the annual premium payable by an individual for reimbursement $j$. The total reimbursement to the group is

$$R = \sum_{i=1}^{m} r[\chi(i), X_i];$$

the total premiums paid are:

$$P = \sum_{i=1}^{m} P[\chi(i)],$$

and

$$G = P - R = \sum_{i=1}^{m} \{P[\chi(i)] - r[\chi(i), X_i]\}$$

is the insurer's gain.

*Example 3.1*

We have a set of $X_i$, $1 \leq i \leq m$, mutually independent and identically distributed as in example 2.10. The set of functions $r_j(x) = r(j,x)$, $1 \leq j \leq n$, is as in example 2.9 where $d(j)$, $c(j)$, $L(j)$ and $M(j)$ correspond to $r_j$, and therefore we have the intervals $B(j)$, $C(j)$ and $D(j)$. For a choice function $\chi$, we can calculate the values:

$$E\{r[\chi(i), X_i]\} = \sum_{ks \in B[\chi(i)]} c[\chi(i)]\{ks - d[\chi(i)]\}p_k + \sum_{ks \in C[\chi(i)]} \{ks - d[\chi(i)] - L[\chi(i)]\}p_k + \sum_{ks \in D[\chi(i)]} M[\chi(i)]p_k$$

and

$$E\{r^2[\chi(i), X_i]\} = \sum_{ks \in B[\chi(i)]} c^2[\chi(i)]\{ks - d[\chi(i)]\}^2 p_k + \sum_{ks \in C[\chi(i)]} \{ks - d[\chi(i)] - L[\chi(i)]\}^2 p_k + \sum_{ks \in D[\chi(i)]} M^2[\chi(i)]p_k.$$

From these we can then (given a set of $P_i$'s) calculate:

$$E[R] = \sum_{i=1}^{m} E\{r[\chi(i), X_i]\},$$

$$\text{Var}[R] = \sum_{i=1}^{m} \text{Var}\{r[\chi(i), X_i]\},$$

$E[G]$, and $\text{Var}[G]$.

*Example 3.2*

We can let the $X_i$ have the distribution of example 2.11. We can also have the reimbursements $r_j$'s and the choice function $\chi(i)$ of example 3.1. Then:

$$E\{r[\chi(i), X_i]\} = \int_{B[\chi(i)]} c[\chi(i)]\{t - d[\chi(i)]\} f(t)\,dt + \int_{C[\chi(i)]} \{t - d[\chi(i)] - L[\chi(i)]\} f(t)\,dt + \int_{D[\chi(i)]} M[\chi(i)] f(t)\,dt$$

and

$$E\{r^2[\chi(i), X_i]\} = \int_{B[\chi(i)]} c^2[\chi(i)]\{t - d[\chi(i)]\}^2 f(t)\,dt + \int_{C[\chi(i)]} \{t - d[\chi(i)] - L[\chi(i)]\}^2 f(t)\,dt + \int_{D[\chi(i)]} M^2[\chi(i)] f(t)\,dt$$

The expressions for $\text{Var}\{r[\chi(i), X_i]\}$, $E[R]$, $\text{Var}[R]$, $E[G]$, and $\text{Var}[G]$ are the same as in example 3.1.

Now we define the "cost deviation due to selection," a random variable for a group with $m$ individuals, as:

$$A = R - \sum_{i=1}^{m} \Psi[\chi(i)]$$
$$= \sum_{i=1}^{m} r[\chi(i), X_i] - \sum_{k=1}^{m} \left[\frac{1}{m} \sum_{i=1}^{m} r[\chi(i), X_k]\right].$$

This is called the cost deviation due to selection because $A$ is equal to the deviation in the reimbursement due to the choice $X$. Since

$$R = A + \sum_{i=1}^{m} \Psi[X(i)]$$

and

$$E[R] = E[A] + E\left[\frac{1}{m}\sum_{i=1}^{m} \Psi[X(i)]\right] = E[A] + \frac{1}{m}\sum_{i=1}^{m} E\{\Psi[X(i)]\},$$

the problem of estimating $E[R]$ is reduced to estimating $E[A]$ and using the traditional rating techniques (for example, manual rates as discussed above) for $E\{\Psi[X(i)]\}$ in the second term.

Here are some of the properties of $A$ (proofs omitted):

I. $A$ is exactly equal to the amount by which the actual reimbursement exceeds what the reimbursement would have been if each individual were reimbursed at the mean rate for the group. That is, if we define the mean reimbursement for the group

$$\bar{r}(x) = \frac{1}{m}\sum_{k=1}^{m} r[X(k), x],$$

then

$$A = \sum_{i=1}^{m} \{r[X(i), X_i] - \bar{r}(X_i)\} = \sum_{i=1}^{m} A(i)$$

for

$$A(i) = r[X(i), X_i] - \bar{r}(X_i).$$

II. If the $X_i$ are identically distributed, then $E(A) = 0$.

III. If $X$ is a constant, $X(1) = X(2) = \ldots = X(m)$, then $A = 0$.

IV. Often the insurer sets $P(i) = E[\Psi(i)]$. In which case $E(G) = -E(A)$.

V. If the values of $X(i)$ are treated as random variables that are independent of the $X_i$, then $E(A) = 0$.

*Example 3.3*

Table 4 presents a hypothetical group with $m = 100$. Shown for each individual is $E(X_i)$ and the choice $X(i)$. Here $n = 4$ and the four choices are 1 through 4 of example 2.9. Table 5 shows the expectations and variances of

TABLE 4

SAMPLE GROUP*

| $i$ | $E[X_i]$ | $x(i)$ | $X_i$ | $r(1, X_i)$ | $r(2, X_i)$ | $r(3, X_i)$ | $r[x(i), X_i]$ | $A(i)$ |
|---|---|---|---|---|---|---|---|---|
| 1... | $ 286.73 | 1 | $ 5 | $ 0 | $ 0 | $ 0 | $ 0 | $ 0 |
| 2... | 286.73 | 1 | 3,358 | 2,858 | 2,287 | 1,769 | 2,858 | 951 |
| 3... | 286.73 | 1 | 4,090 | 3,590 | 2,872 | 2,317 | 3,590 | 1,158 |
| 4... | 286.73 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5... | 286.73 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6... | 286.73 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7... | 286.73 | 3 | 478 | 302 | 0 | 0 | 0 | (73) |
| 8... | 286.73 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9... | 286.73 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10... | 286.73 | 4 | 0 | 0 | 0 | 0 | 0 | 0 |
| 11... | 286.73 | 4 | 0 | 0 | 0 | 0 | 0 | 0 |
| 12... | 286.73 | 4 | 0 | 0 | 0 | 0 | 0 | 0 |
| 13... | 286.73 | 4 | 0 | 0 | 0 | 0 | 0 | 0 |
| 14... | 645.15 | 3 | 1,000 | 720 | 400 | 0 | 0 | (269) |
| 15... | 645.15 | 3 | 1,522 | 1,138 | 818 | 392 | 392 | (226) |
| 16... | 645.15 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 17... | 645.15 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 18... | 645.15 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 19... | 645.15 | 3 | 1,211 | 889 | 569 | 158 | 158 | (252) |
| 20... | 645.15 | 3 | 707 | 486 | 166 | 0 | 0 | (156) |
| 21... | 645.15 | 4 | 102 | 2 | 0 | 0 | 0 | 0 |
| 22... | 1,146.94 | 3 | 512 | 330 | 10 | 0 | 0 | (81) |
| 23... | 1,146.94 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 24... | 1,146.94 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 25... | 1,146.94 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 26... | 1,146.94 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 27... | 1,146.94 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 28... | 1,146.94 | 4 | 551 | 360 | 40 | 0 | 0 | (96) |
| 29... | 1,577.04 | 1 | 2,115 | 1,615 | 1,292 | 836 | 1,615 | 600 |
| 30... | 1,577.04 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 31... | 1,577.04 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 32... | 1,577.04 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 33... | 1,577.04 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 34... | 1,577.04 | 3 | 1,798 | 1,359 | 1,039 | 599 | 599 | (204) |
| 35... | 1,863.78 | 1 | 15,396 | 14,896 | 13,896 | 11,396 | 14,896 | 3,655 |
| 36... | 1,863.78 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 37... | 1,863.78 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 38... | 1,863.78 | 3 | 213 | 90 | 0 | 0 | 0 | (22) |
| 39... | 1,863.78 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 40... | 1,863.78 | 3 | 295 | 156 | 0 | 0 | 0 | (37) |
| 41... | 2,293.88 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 42... | 2,293.88 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 43... | 2,293.88 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 44... | 2,293.88 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 45... | 2,293.88 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 46... | 2,293.88 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 47... | 3,154.08 | 1 | 5,795 | 5,295 | 4,295 | 3,596 | 5,295 | 1,627 |
| 48... | 3,154.08 | 1 | 6,588 | 6,088 | 5,088 | 4,191 | 6,088 | 1,813 |
| 49... | 3,154.08 | 1 | 15,649 | 15,149 | 14,149 | 11,649 | 15,149 | 3,691 |
| 50... | 3,154.08 | 1 | 39,806 | 39,306 | 38,306 | 35,806 | 39,306 | 7,073 |

TABLE 4–*Continued*

| $i$ | $E[X_i]$ | $x(i)$ | $X_i$ | $r(1, X_i)$ | $r(2, X_i)$ | $r(3, X_i)$ | $r[x(i), X_i]$ | $A(i)$ |
|---|---|---|---|---|---|---|---|---|
| 51... | 3,154.08 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 52... | 4,014.29 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 53... | 4,014.29 | 1 | 593 | 394 | 74 | 0 | 394 | 282 |
| 54... | 4,014.29 | 1 | 4,960 | 4,460 | 3,568 | 2,970 | 4,460 | 1,405 |
| 55... | 4,014.29 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 56... | 573.47 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 57... | 573.47 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 58... | 573.47 | 2 | 1,084 | 787 | 467 | 63 | 467 | 142 |
| 59... | 573.47 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 60... | 573.47 | 3 | 794 | 555 | 235 | 0 | 0 | (190) |
| 61... | 573.47 | 3 | 1,104 | 803 | 483 | 78 | 78 | (260) |
| 62... | 573.47 | 3 | 275 | 140 | 0 | 0 | 0 | (34) |
| 63... | 573.47 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 64... | 573.47 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 65... | 573.47 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 66... | 573.47 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 67... | 573.47 | 4 | 0 | 0 | 0 | 0 | 0 | 0 |
| 68... | 573.47 | 4 | 0 | 0 | 0 | 0 | 0 | 0 |
| 69... | 573.47 | 4 | 39 | 0 | 0 | 0 | 0 | 0 |
| 70... | 573.47 | 4 | 0 | 0 | 0 | 0 | 0 | 0 |
| 71... | 573.47 | 4 | 0 | 0 | 0 | 0 | 0 | 0 |
| 72... | 573.47 | 4 | 0 | 0 | 0 | 0 | 0 | 0 |
| 73... | 1,003.57 | 1 | 1,891 | 1,433 | 1,113 | 668 | 1,433 | 568 |
| 74... | 1,003.57 | 2 | 1,780 | 1,344 | 1,024 | 585 | 1,024 | 233 |
| 75... | 1,003.57 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 76... | 1,003.57 | 3 | 965 | 692 | 372 | 0 | 0 | (256) |
| 77... | 1,003.57 | 3 | 2,261 | 1,761 | 1,409 | 946 | 946 | (174) |
| 78... | 1,003.57 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 79... | 1,003.57 | 4 | 0 | 0 | 0 | 0 | 0 | 0 |
| 80... | 1,003.57 | 4 | 0 | 0 | 0 | 0 | 0 | 0 |
| 81... | 1,146.94 | 1 | 5,563 | 5,063 | 4,063 | 3,422 | 5,063 | 1,572 |
| 82... | 1,146.94 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 83... | 1,146.94 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 84... | 1,146.94 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 85... | 1,146.94 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 86... | 1,146.94 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 87... | 1,146.94 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 88... | 2,007.14 | 1 | 7,311 | 6,811 | 5,811 | 4,733 | 6,811 | 1,983 |
| 89... | 2,007.14 | 2 | 997 | 717 | 397 | 0 | 397 | 130 |
| 90... | 2,007.14 | 2 | 1,218 | 895 | 575 | 164 | 575 | 160 |
| 91... | 2,007.14 | 2 | 4,536 | 4,036 | 3,229 | 2,652 | 3,229 | 477 |
| 92... | 2,007.14 | 2 | 232 | 106 | 0 | 0 | 0 | (25) |
| 93... | 2,437.25 | 1 | 1,883 | 1,426 | 1,106 | 662 | 1,426 | 567 |
| 94... | 2,437.25 | 1 | 3,754 | 3,254 | 2,603 | 2,066 | 3,254 | 1,063 |
| 95... | 2,437.25 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 96... | 2,437.25 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 97... | 2,867.35 | 1 | 6,751 | 6,251 | 5,251 | 4,313 | 6,251 | 1,851 |
| 98... | 2,867.35 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 99... | 3,297.45 | 1 | 2,708 | 2,208 | 1,767 | 1,281 | 2,208 | 767 |
| 100... | 3,584.19 | 1 | 2,079 | 1,583 | 1,263 | 809 | 1,583 | 593 |
| Total | $141,360 | | $153,970 | $139,349 | $120,037 | $98,122 | $129,546 | $30,007 |

*Number Selecting Reimbursements:

| $j$ | no. |
|---|---|
| 1 | 24 |
| 2 | 24 |
| 3 | 38 |
| 4 | 14 |

TABLE 5

EXPECTATION, VARIANCE, AND STANDARD DEVIATION
OF MEAN REIMBURSEMENTS *R/m* AND *A/m*
SAMPLE SELECTION AND DISTRIBUTIONS BASED ON UNADJUSTED EXPECTED VALUES

| Reimbursement Number (*r*) | Number Selecting | Discrete Distribution | | | Pareto Distribution | | |
|---|---|---|---|---|---|---|---|
| | | Mean | Variance | Standard Deviation | Mean | Variance | Standard Deviation |
| 1 | 24 | $1,871 | 818,820 | $905 | $2,764 | 1,301,139 | $1,141 |
| 2 | 24 | 1,668 | 796,109 | 892 | 2,457 | 1,284,966 | 1,134 |
| 3 | 38 | 1,411 | 693,181 | 833 | 2,083 | 1,008,308 | 1,004 |
| 4 | 14 | 0 | 0 | 0 | 0 | 0 | 0 |
| Covered Charges | | 2,027 | 851,073 | 923 | 3,021 | 3,407,162 | 1,846 |
| *R/m* | | 1,564 | 774,686 | 880 | 2,472 | 1,272,618 | 1,128 |
| *A/m* | | 178.543 | 33,655 | 183 | 457.442 | 53,851 | 232 |

$\Psi(j)$ $(1 \leq j \leq 4)$, $R/m$, and $A/m$. These have been calculated under two assumptions: (1) each $X_i$ has the distribution of example 2.10 with $s = E(X_i)/1433.67$, and (2) each $X_i$ has the distribution of example 2.11 (Table 3, Pareto) with $\lambda = E(X_i)(1.15738)$. This value of $\lambda$ will give a Pareto distribution with the required expectation.

Table 4 also shows, for each individual in the group, an example outcome of values for $X_i$, the corresponding values of $r(j,X_i)$ for $j = 1$, 2, and 3, and the value of $A(i)$. Thus there were covered charges of $153,970 (compared to the expected value of $141,360), reimbursements $R$ of $129,546, and $A$ of $30,007.

The values of $E(X_i)$ can be thought of as the expected covered charges due to known (to the insurer) characteristics of the individuals in the group, such as their ages. In such case $E(A)$ can be thought of as the expected cost deviation due to demographic selection. If the actual value of $A$ greatly exceeds this $E(A)$, then the insurer might wonder whether the individuals knew more about their health status and used this knowledge to antiselect. We can approximate the probability that a value of $A$ was realized randomly by using $E(A)$ and Var$(A)$ with the normal approximation.

## IV. PRIOR-YEAR'S CHARGES

Let us assume that each individual has a, possibly unknown, parameter for the distribution of his or her covered charges. We call this parameter $y = \{y(i) | 1 \leq i \leq m\}$, where $y(i)$ pertains to individual $i$. Note that the $y(i)$'s could themselves be treated as realizations of random variables $Y(i)$'s and may be multidimensional. In any case, if we knew the values of the $y(i)$'s, we could calculate $E[A|y]$. Since there is generally a correlation between successive years' charges, we could take a set of $y(i)$'s to be each individual's prior-year's charges.[4]

*Example 4.1*

Table 6 expands Table 4. The values that were previously called $X_i$ are now taken to represent last year's claims and are identified as $y(i)$. Table 6 also shows a value of $E[X_i|Y_i = y(i)]$. Here we have set $E[X_i|y(i)] = 0.75E[X_i] + 0.25y(i)$. Table 7 shows the $E[\Psi(j)|y]$, $(1 \leq j \leq 4)$, $\text{Var}[\Psi(j)|y]$, $E[R/m|y]$, $\text{Var}[R/m|y]$, $E[A/m|y]$ and $\text{Var}[A/m|y]$. These are computed using the two assumptions of example 3.3. We have assumed that the $X_i$ always have the same distributions except for a scale change.

*Example 4.2*

Very often the parameter $y$ would be unknown. If we assume that it is equal to the prior-year's charges, we could assume that each $y(i)$ has the distribution of $X_i$. If we set $E[X_i|Y_i = y(i)] = 0.75E[X_i] + 0.25y(i)$, then we can calculate $E[R] = E[E(R|Y)]$ and $\text{Var}[R] = \text{Var}[E(R|Y)] + E[\text{Var}(R|Y)]$. The calculations involved are long and tedious, so no example values have been calculated. A Monte Carlo simulation technique could be used instead.

## V. PREDICTING CHOICE

To predict employee choice, we assume that each of the individuals, $i(1 \leq i \leq m)$ has a utility function $u_i(w)$ for wealth $w \geq 0$.[5] Now we assume that each individual will select the reimbursement that maximizes his or her expected utility. That is, if each individual's initial wealth is $w(i)$ and there exists a $1 \leq k \leq n$ such that:

$$E\{u_i[w(i) - X_i + r(k,X_i) - P(k)]\} \geq E\{u_i[w(i) - X_i + r(j,X_i) - P(j)]\}$$

for every $j$, $1 \leq j \leq n$, then $X(i) = k$. Trivially, if there are two (or more) reimbursements for which the expected utility is equal and greater than all the other reimbursements, we will assume an arbitrary selection.

[4]Fuhrer [12, p. 403] found a correlation of 24.35 percent, and Cookson [8, p. 1602] reported seeing estimates of 15 to 25 percent.

[5]See [6, Chapter 1] for an introduction to risk-averse utility functions. A good reference on utility functions is [16], particularly Chapter 4, which has an excellent section on various types of utility functions.

TABLE 6

SAMPLE GROUP*

| $i$ | $E[X_i]$ | $x(i)$ | $X_i = y(i)$ | $E[X_i \mid y(i)]$ | $x(i)$ (1) | $x(i)$ (2) | $x(i)$ (3) | $a(i)$ |
|---|---|---|---|---|---|---|---|---|
| 1... | $ 286.73 | 1 | $ 5 | $ 216 | 3 | 3 | 3 | $ 6,000 |
| 2... | 286.73 | 1 | 3,358 | 1,055 | 3 | 1 | 2 | 8,500 |
| 3... | 286.73 | 1 | 4,090 | 1,238 | 2 | 1 | 2 | 3,600 |
| 4... | 286.73 | 2 | 0 | 215 | 2 | 2 | 2 | 3,400 |
| 5... | 286.73 | 2 | 0 | 215 | 3 | 3 | 3 | 6,800 |
| 6... | 286.73 | 2 | 0 | 215 | 3 | 3 | 3 | 7,600 |
| 7... | 286.73 | 3 | 478 | 335 | 3 | 3 | 3 | 7,500 |
| 8... | 286.73 | 3 | 0 | 215 | 2 | 2 | 3 | 4,400 |
| 9... | 286.73 | 3 | 0 | 215 | 2 | 2 | 3 | 4,900 |
| 10... | 286.73 | 4 | 0 | 215 | 3 | 3 | 3 | 7,200 |
| 11... | 286.73 | 4 | 0 | 215 | 3 | 3 | 3 | 6,500 |
| 12... | 286.73 | 4 | 0 | 215 | 2 | 2 | 3 | 4,100 |
| 13... | 286.73 | 4 | 0 | 215 | 3 | 3 | 3 | 8,500 |
| 14... | 645.15 | 3 | 1,000 | 734 | 3 | 3 | 2 | 11,000 |
| 15... | 645.15 | 3 | 1,522 | 864 | 3 | 3 | 2 | 14,500 |
| 16... | 645.15 | 3 | 0 | 484 | 3 | 3 | 3 | 9,500 |
| 17... | 645.15 | 3 | 0 | 484 | 3 | 3 | 3 | 13,100 |
| 18... | 645.15 | 3 | 0 | 484 | 2 | 2 | 3 | 5,400 |
| 19... | 645.15 | 3 | 1,211 | 787 | 3 | 3 | 2 | 12,300 |
| 20... | 645.15 | 3 | 707 | 661 | 3 | 3 | 3 | 9,000 |
| 21... | 645.15 | 4 | 102 | 509 | 1 | 2 | 2 | 3,800 |
| 22... | 1,146.94 | 3 | 512 | 988 | 1 | 2 | 2 | 19,800 |
| 23... | 1,146.94 | 3 | 0 | 860 | 1 | 3 | 2 | 23,500 |
| 24... | 1,146.94 | 3 | 0 | 860 | 1 | 2 | 2 | 6,500 |
| 25... | 1,146.94 | 3 | 0 | 860 | 1 | 3 | 2 | 19,600 |
| 26... | 1,146.94 | 3 | 0 | 860 | 1 | 3 | 2 | 17,500 |
| 27... | 1,146.94 | 3 | 0 | 860 | 1 | 3 | 2 | 20,900 |
| 28... | 1,146.94 | 4 | 551 | 998 | 1 | 2 | 2 | 14,400 |
| 29... | 1,577.04 | 1 | 2,115 | 1,712 | 1 | 1 | 2 | 31,800 |
| 30... | 1,577.04 | 2 | 0 | 1,183 | 1 | 1 | 2 | 20,100 |
| 31... | 1,577.04 | 2 | 0 | 1,183 | 1 | 1 | 2 | 30,300 |
| 32... | 1,577.04 | 2 | 0 | 1,183 | 1 | 1 | 2 | 31,400 |
| 33... | 1,577.04 | 3 | 0 | 1,183 | 1 | 1 | 2 | 25,200 |
| 34... | 1,577.04 | 3 | 1,798 | 1,632 | 1 | 1 | 2 | 7,700 |
| 35... | 1,863.78 | 1 | 15,396 | 5,247 | 1 | 1 | 2 | 11,300 |
| 36... | 1,863.78 | 2 | 0 | 1,398 | 1 | 1 | 2 | 28,400 |
| 37... | 1,863.78 | 2 | 0 | 1,398 | 1 | 1 | 2 | 36,000 |
| 38... | 1,863.78 | 3 | 213 | 1,451 | 1 | 1 | 2 | 26,600 |
| 39... | 1,863.78 | 3 | 0 | 1,398 | 1 | 1 | 2 | 7,800 |
| 40... | 1,863.78 | 3 | 295 | 1,472 | 1 | 1 | 2 | 21,700 |
| 41... | 2,293.88 | 1 | 0 | 1,720 | 1 | 1 | 2 | 35,000 |
| 42... | 2,293.88 | 1 | 0 | 1,720 | 1 | 1 | 2 | 15,900 |
| 43... | 2,293.88 | 1 | 0 | 1,720 | 1 | 1 | 2 | 27,600 |
| 44... | 2,293.88 | 2 | 0 | 1,720 | 1 | 1 | 2 | 20,800 |
| 45... | 2,293.88 | 2 | 0 | 1,720 | 1 | 1 | 2 | 39,300 |
| 46... | 2,293.88 | 3 | 0 | 1,720 | 1 | 1 | 2 | 45,200 |
| 47... | 3,154.08 | 1 | 5,795 | 3,814 | 1 | 1 | 2 | 17,400 |
| 48... | 3,154.08 | 1 | 6,588 | 4,013 | 1 | 1 | 2 | 10,800 |
| 49... | 3,154.08 | 1 | 15,649 | 6,278 | 1 | 1 | 2 | 59,900 |
| 50... | 3,154.08 | 1 | 39,806 | 12,317 | 1 | 1 | 2 | 8,200 |

TABLE 6—*Continued*

| $i$ | $E[X_i]$ | $x(i)$ | $X_i = y(i)$ | $E[X_i \mid y(i)]$ | $x(i)$ (1) | (2) | (3) | $a(i)$ |
|---|---|---|---|---|---|---|---|---|
| 51... | 3,154.08 | 2 | 0 | 2,366 | 1 | 1 | 2 | 49,800 |
| 52... | 4,014.29 | 1 | 0 | 3,011 | 1 | 1 | 2 | 6,500 |
| 53... | 4,014.29 | 1 | 593 | 3,159 | 1 | 1 | 2 | 79,900 |
| 54... | 4,014.29 | 1 | 4,960 | 4,251 | 1 | 1 | 2 | 10,000 |
| 55... | 4,014.29 | 2 | 0 | 3,011 | 1 | 1 | 2 | 40,200 |
| 56... | 573.47 | 1 | 0 | 430 | 3 | 3 | 3 | 9,500 |
| 57... | 573.47 | 2 | 0 | 430 | 3 | 3 | 3 | 9,900 |
| 58... | 573.47 | 2 | 1,084 | 701 | 3 | 3 | 2 | 8,300 |
| 59... | 573.47 | 2 | 0 | 430 | 3 | 3 | 3 | 8,300 |
| 60... | 573.47 | 3 | 794 | 629 | 3 | 3 | 3 | 14,300 |
| 61... | 573.47 | 3 | 1,104 | 706 | 3 | 3 | 2 | 7,400 |
| 62... | 573.47 | 3 | 275 | 499 | 3 | 3 | 3 | 6,900 |
| 63... | 573.47 | 3 | 0 | 430 | 3 | 3 | 3 | 14,200 |
| 64... | 573.47 | 3 | 0 | 430 | 3 | 3 | 3 | 6,300 |
| 65... | 573.47 | 3 | 0 | 430 | 1 | 2 | 2 | 3,800 |
| 66... | 573.47 | 3 | 0 | 430 | 3 | 3 | 3 | 12,600 |
| 67... | 573.47 | 4 | 0 | 430 | 3 | 3 | 3 | 9,400 |
| 68... | 573.47 | 4 | 0 | 430 | 2 | 2 | 3 | 4,700 |
| 69... | 573.47 | 4 | 39 | 440 | 3 | 3 | 3 | 13,400 |
| 70... | 573.47 | 4 | 0 | 430 | 3 | 3 | 3 | 8,500 |
| 71... | 573.47 | 4 | 0 | 430 | 3 | 3 | 3 | 7,600 |
| 72... | 573.47 | 4 | 0 | 430 | 3 | 3 | 3 | 5,700 |
| 73... | 1,003.57 | 1 | 1,891 | 1,225 | 2 | 1 | 2 | 14,700 |
| 74... | 1,003.57 | 2 | 1,780 | 1,198 | 1 | 1 | 2 | 5,900 |
| 75... | 1,003.57 | 3 | 0 | 753 | 1 | 3 | 2 | 8,700 |
| 76... | 1,003.57 | 3 | 965 | 994 | 2 | 2 | 2 | 13,900 |
| 77... | 1,003.57 | 3 | 2,261 | 1,318 | 2 | 1 | 2 | 13,000 |
| 78... | 1,003.57 | 3 | 0 | 753 | 1 | 3 | 2 | 8,400 |
| 79... | 1,003.57 | 4 | 0 | 753 | 1 | 1 | 2 | 5,000 |
| 80... | 1,003.57 | 4 | 0 | 753 | 2 | 3 | 2 | 12,300 |
| 81... | 1,146.94 | 1 | 5,563 | 2,251 | 1 | 1 | 2 | 12,800 |
| 82... | 1,146.94 | 2 | 0 | 860 | 1 | 1 | 2 | 4,700 |
| 83... | 1,146.94 | 2 | 0 | 860 | 1 | 1 | 2 | 4,300 |
| 84... | 1,146.94 | 3 | 0 | 860 | 1 | 3 | 2 | 16,300 |
| 85... | 1,146.94 | 3 | 0 | 860 | 1 | 3 | 2 | 16,700 |
| 86... | 1,146.94 | 3 | 0 | 860 | 1 | 2 | 2 | 6,500 |
| 87... | 1,146.94 | 3 | 0 | 860 | 1 | 3 | 2 | 23,900 |
| 88... | 2,007.14 | 1 | 7,311 | 3,333 | 1 | 1 | 2 | 28,900 |
| 89... | 2,007.14 | 2 | 997 | 1,755 | 1 | 1 | 2 | 4,200 |
| 90... | 2,007.14 | 2 | 1,218 | 1,810 | 1 | 1 | 2 | 31,600 |
| 91... | 2,007.14 | 2 | 4,536 | 2,639 | 1 | 1 | 2 | 34,200 |
| 92... | 2,007.14 | 2 | 232 | 1,563 | 1 | 1 | 2 | 9,000 |
| 93... | 2,437.25 | 1 | 1,883 | 2,299 | 1 | 1 | 2 | 24,700 |
| 94... | 2,437.25 | 1 | 3,754 | 2,766 | 1 | 1 | 2 | 14,600 |
| 95... | 2,437.25 | 2 | 0 | 1,828 | 1 | 1 | 2 | 28,400 |
| 96... | 2,437.25 | 3 | 0 | 1,828 | 1 | 1 | 2 | 18,200 |
| 97... | 2,867.35 | 1 | 6,751 | 3,838 | 1 | 1 | 2 | 10,400 |
| 98... | 2,867.35 | 2 | 0 | 2,151 | 1 | 1 | 2 | 27,700 |
| 99... | 3,297.45 | 1 | 2,708 | 3,150 | 1 | 1 | 2 | 47,400 |
| 100... | 3,584.19 | 1 | 2,079 | 3,208 | 1 | 1 | 2 | 45,100 |
| Total.. | $141,360 | | $153,970 | $139,349 | | | | |

*Number Selecting Reimbursements:

| $j$ | Sample | (1) | (2) | (3) |
|---|---|---|---|---|
| 1 | 24 | 60 | 49 | 0 |
| 2 | 24 | 11 | 13 | 72 |
| 3 | 38 | 29 | 38 | 28 |
| 4 | 14 | 0 | 0 | 0 |

TABLE 7

EXPECTATION, VARIANCE, AND STANDARD DEVIATIONS OF MEAN REIMBURSEMENTS, $R/m$, AND $A/m$
SAMPLE SELECTION AND DISTRIBUTIONS BASED ON Y CONDITIONED EXPECTED VALUES

| Reimbursement Number ($r$) | Number Selecting | Discrete Distribution Mean | Discrete Distribution Variance | Discrete Distribution Standard Deviation | Pareto Distribution Mean | Pareto Distribution Variance | Pareto Distribution Standard Deviation |
|---|---|---|---|---|---|---|---|
| 1 | 24 | $1,919 | 1,134,917 | $1,065 | $4,178 | 5,435,419 | $2,331 |
| 2 | 24 | 1,721 | 1,112,045 | 1,055 | 3,892 | 5,423,544 | 2,329 |
| 3 | 38 | 1,465 | 902,406 | 950 | 3,366 | 3,243,718 | 1,801 |
| 4 | 14 | 0 | 0 | 0 | 0 | 0 | 0 |
| Covered Charges | | 2,072 | 1,295,754 | 1,138 | 4,599 | 37,442,838 | 6,119 |
| $R/m$ | | 1,657 | 1,104,684 | 1,051 | 3,959 | 5,421,039 | 2,328 |
| $A/m$ | | 226.304 | 48,116 | 219 | 742.948 | 359,771 | 600 |

For simplicity, we want to use the same form of a utility function for each individual. To model the actual situation, we need to say that each individual has a different aversion to risk. To do this, we select a utility function that is decreasingly risk-averse. That is, the larger the individual's initial wealth, the less risk-averse he or she is. Common measures of risk aversion are the Arrow-Pratt [2] and [21] measures of absolute risk aversion and relative risk aversion: $\rho_u(w) = -u''(w)/u'(w)$ and $\delta_u(w) = w\rho_u(w)$, respectively.[6]

*Example 5.1*

We can use the assumptions of example 3.3 in which the choice depends on the utility function: $u_i(w) = \ln[w + a(i)]$ for a positive constant $a(i)$. This utility function is convenient because of the property that almost any level of risk averseness can be selected based on the size of the parameter $a(i)$.[7] Table 6 shows some sample values of $a(i)$ for our sample group and the resulting choice in column 1 using the discrete distribution to calculate expectations. Note that we have changed slightly the reimbursements so that there is no maximum $M$. The end of Table 6 summarizes the choices, and Table 8 shows the calculated values. We have assumed that $P(j) = E[\Psi(j)]$.

[6]Kimball [17, p. 2] suggests "standard risk aversion" as another alternative. It is characteristic of utility functions associated with constant relative risk aversion.

[7]For $u(w) = \ln(w)$, the absolute risk aversion is $\rho_u(w) = 1/w$, which is a decreasing function of $w$, and the relative risk aversion is $\delta_u(w) = 1$.

TABLE 8

VALUES FOR $R/m$ AND $A/m$

| Example | | Discrete Distribution Mean | Discrete Distribution Variance | Discrete Distribution Standard Deviation | Pareto Distribution Mean | Pareto Distribution Variance | Pareto Distribution Standard Deviation |
|---|---|---|---|---|---|---|---|
| 5.1 | $R/m$ | \$1,779 | 812,894 | \$ 902 | \$2,715 | 1,300,147 | \$1,140 |
| | $A/m$ | 63 | 2,011 | 45 | 182 | 5,907 | 77 |
| 5.2 | $R/m$ | 1,831 | 1,129,125 | 1,063 | 4,129 | 5,434,103 | 2,331 |
| | $A/m$ | 65 | 4,412 | 66 | 218 | 58,528 | 242 |
| 5.3 | $R/m$ | 1,795 | 1,126,295 | 1,061 | 4,104 | 5,433,383 | 2,331 |
| | $A/m$ | 74 | 7,302 | 85 | 272 | 100,422 | 317 |
| 5.5 | $R/m$ | 1,697 | 1,110,082 | 1,054 | 3,887 | 5,423,368 | 2,329 |
| | $A/m$ | 48 | 3,605 | 60 | 142 | 54,459 | 233 |

*Example 5.2*

For this example, use the assumptions of example 4.1, with a fixed known parameter set $y(i)$ and with the utility-based choice of example 5.1. The calculated values are also shown in Table 8.

*Example 5.3*

This is example 5.1, except we use the parameter-adjusted discrete distribution of example 4.1 to calculate the expected utilities and determine the choices. Table 6 shows the choices in column 2, and Table 8 shows the calculated values using the parameter-adjusted distributions as in example 5.2. Note that choice 2 has a larger $E[A]$ than choice 1.

*Example 5.4*

Here we combine example 4.2 with the utility function of example 5.1. Now that the choice is random, we could calculate, for each $i$ and $j$, $Pr\{X(i)=j\}$. We define $N(j)$ as the number of individuals for whom $X(i)=j$. We could also calculate $E[N(j)]$, $1 \leq j \leq 4$.

*Example 5.5*

Let $S(j)=\{i:X(i)=j\}$. Then let

$$P(j) = \frac{1}{N(j)} \sum_{i \in S(j)} r[y(i)]$$

in example 5.2. That is, we set the premiums for a reimbursement equal to the experience of those who selected it (using the sample selection). The resulting choice (Table 6, column 3) is much more heavily weighted towards the cheaper plans. This illustrates the selection spiral that can occur if premium rates are based only on the experience of those who choose a particular reimbursement plan.

## VI. CONCLUSION AND AREAS FOR FURTHER RESEARCH

The framework of this paper allows us to predict employee choice and cost deviations due to selection given any arbitrary combination of individual charge distributions, a set of reimbursement plans and their premiums, and a set of utility functions. By using this method, various combinations of plans and premiums can be explored until the plan administrator can select the combination that best fits the group's needs.

The calculations of examples 4.2 and 5.4 could be completed. A few more distributions could be used to calculate the values. A term could be added to each reimbursement's wealth to model affinities that individuals may have for a particular plan. This might be used in the HMO choice, as individuals might prefer the traditional plan over the HMO so that they could continue with their current physicians.

The parameters of the utility function could be estimated from some actual choice data. These could then be used to predict actual past choices and then to determine the accuracy of the predictions.

## REFERENCES

1. AKERLOF, G.A. "The Market for Lemons: Quality, Uncertainty, and the Market Mechanism," *Quarterly Journal of Economics* 84 (1970): 488–500.
2. ARROW, K.J. "The Theory of Risk Aversion," Lecture 2 in *Aspects of the Theory of Risk Bearing*, Yrjö Jahnssonin Säätio, Helsinki; Reprinted in *Collected Papers of Kenneth J. Arrow, Vol 3: Individual Choice under Certainty and Uncertainty*. Cambridge: The Belnap Press of Harvard University Press, 1984.
3. ARROW, K.J. *Essays in the Theory of Risk Bearing*. Amsterdam: North-Holland Publishing Co., 1971.
4. BLAZENKO, G. "Optimal Indemnity Contracts," *Insurance: Mathematics and Economics* 4, no. 4 (October 1985): 267–78.
5. BORCH, K. "Equilibrium in a Reinsurance Market," *Econometrica* 30 (1962): 424–44.
6. BOWERS, N.L., Jr., GERBER, H.U., HICKMAN, J.C., JONES, D.A., AND NESBITT, C.J. *Actuarial Mathematics*. Chicago, Ill.: Society of Actuaries, 1986.
7. BÜHLMANN, H., AND JEWELL, W. "Optimal Risk Exchanges," *ASTIN Bulletin* 10, no. 3 (1979): 243–62.
8. COOKSON, J.P. "High Cost Claimants in Medical Benefit Plans," *RSA* 15 (1989): 1601–4.

9. DEGROOT, M. *Optimal Statistical Decisions*. New York: McGraw-Hill, 1970.
10. DOHERTY, N.A., AND SCHLESINGER, H. "Optimal Insurance in Incomplete Markets," *Journal of Political Economy* 91 (1990): 1045–54.
11. DOHERTY, N.A., AND SCHLESINGER, H. "Rational Insurance Purchasing: Consideration of Contract Nonperformance," *Quarterly Journal of Economics* 105 (1990): 243–53.
12. FUHRER, C.S. "Some Applications of Credibility Theory to Group Insurance," *TSA* XL, Pt. 1 (1988): 387–404.
13. GOULD, J.P. "The Expected Utility Hypothesis and the Selection of Optimal Deductibles for a Given Insurance Policy," *Journal of Business* 42, no. 2 (1969): 143–51.
14. HIGHAN, J.A. "On the Value of Selection as Exercised by the Policy-holder against the Company," *Journal of the Institute* I (1851): 179–202.
15. HOGG, R., AND KLUGMAN, S. *Loss Distributions*. New York: John Wiley and Sons, 1984.
16. KEENEY, R.L., AND RAIFFA, H. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. New York: John Wiley and Sons, 1976.
17. KIMBALL, M. "Standard Risk Aversion," working paper, University of Michigan, 1989.
18. MAYERS, D., AND SMITH, C.W., JR. "The Interdependence of Individual Portfolio Decisions and the Demand for Insurance," *Journal of Political Economy* 91, no. 2 (1983): 304–11.
19. MCCLINTOCK, E. "On the Effect of Selection," *TASA* III (1892): 59.
20. MIYAZAKI, H. "The Rat Race and Internal Labor Markets," *The Bell Journal of Economics* 8 (1977): 394–418.
21. PRATT, J.W. "Risk Aversion in the Small and the Large," *Econometrica* 32 (1964): 122–36.
22. RAVIV, A. "The Design of an Optimal Insurance Policy," *American Economic Review* 69 (1979): 84–96.
23. ROTHCHILD, M., AND STIGLITZ, J. "Equilibrium in Insurance Markets: An Essay on the Economics of Imperfect Information," *Quarterly Journal of Economics* 90 (1976): 629–49.
24. SPENCE, M. "Product Differentiation and Performance in Insurance Markets," *Journal of Public Economics* 10 (1978): 427–47.
25. WILSON, C. "A Model of Insurance Markets with Incomplete Information," *Journal of Economic Theory* 97 (1977): 167–207.

## DISCUSSION OF PRECEDING PAPER

GERARD SMEDINGHOFF:

Mr. Fuhrer and Dr. Shapiro present a valuable comprehensive model in an emerging field of research: measuring the causes and effects of individual employee health-care benefit selection. Because this model represents an initial effort in this area, actuaries will eventually see the need to modify it to suit their needs and their own emerging experience. Two refinements and one enhancement are suggested below.

The first refinement concerns how the model portrays employees who opt out of the plan. In the Table 4 example, the values of $A(i)$, "the cost deviation due to selection," are calculated as $r[\chi(i), X_i]-\bar{r}(X_i)$, the mean reimbursement function for the group. The derivation of $\bar{r}(X_i)$ assumes that 14 percent of eligible employees choose a zero reimbursement function (that is, no health coverage). In reality, most of those who decline health coverage are actually choosing to purchase coverage under their spouse's plan as opposed to (a) not wanting the product, (b) viewing the cost as outside of the valid range of their utility curves, or (c) having an array of other vendors to choose from.

Ideally, the model should differentiate between employees who choose no coverage versus those who choose to be covered under their spouse's plan. Because the latter group *does* want health coverage, they *are* willing and able to pay for it, and the coverage available from their spouse's plan may cease or become relatively more expensive, these employees *will* return to the plan since their employer represents the only available vendor of health care coverage. If the employees who opt out were ignored in the calculation of the mean reimbursement function, $\bar{r}(X_i)$, then the resulting value of $A$ in the Table 4 example would be \$13,744 instead of \$30,007.

Actually accounting for this differentiation may not be realistic. From the employer's perspective for the current plan year, the claims generated by employees who opt out of the plan are of no concern. But actuaries should at least note that any changes in the benefit composition, the options offered or the employer contribution rates could significantly affect the size and demographic composition of the percentage of employees who opt out of the plan by either attracting dependents of employees into the plan or driving employees away to be covered as dependents under their spouses' plans.

The second refinement, related to the first, concerns the individual values of $A(i)$ for employees who generate large claims. Almost one-half of the total value of $A$ ($30,007) in Table 4 is attributed to the three employees who generated claims of more than $10,000. As reimbursements to these employees progress to catastrophic levels, their corresponding values of $A(i)$ continue to increase, even after they exceed the maximum out-of-pocket differential (again, this is due to the 14 percent of the mean reimbursement for the group, $\bar{r}(x)$, who are assumed to choose no coverage). Because the employer must pay the bulk of these claims regardless of which reimbursement option is selected, the cost deviation value for any one employee should be limited to the out-of-pocket difference between the highest and lowest non-zero reimbursement options (much in the same way that claims for an individual above a specified level are pooled and not charged to a group's experience). In the Table 4 example, this would reduce the value of $A$ from $30,007 to $26,088.

These two refinements depend on the demographic composition of the subgroup of employees who opt out of the plan. If they appear to be younger, healthier lives who are not representative of the group as a whole, then the refinements may be justified. If there is no discernible pattern to those who opt out, then it appears that they probably represent their alloted portion of claims among all the employees (that is, if, in the example, the 14 percent who opt out also represent 14 percent of the total potential for claims among all employees), then these refinements would not be necessary.

The enhancement concerns the structure of this model and its applications in Sections IV and V of the paper, which suggest a game theory approach to employee benefit selection. In addition to factoring in an employee's wealth and prior year's claims, the model also could consider an employee's deduction into a flexible spending account (FSA) as a measure of the degree of confidence the employee has in the level of health-care costs for the coming plan year. If the FSA deduction were to equal or exceed the chosen reimbursement option's out-of-pocket limit, then that would indicate that the employee may have already planned on significant health-care costs at the time of the annual benefit selection.

Depending on the degree of precision desired, with respect to the employee utility function in Section V, the premiums, $P(j)$, could be adjusted to the after-tax equivalent that would be available as discretionary income to employees who opt out or who choose lower premium options. And if an FSA is available from which employees can fund their out-of-pocket health-care expenses on a pre-tax basis, then the difference between expenses incurred and the amount reimbursed, $X_i - r(j, X_i)$, could also be adjusted to its after-tax equivalent as it would affect an employee's wealth.

The impetus for this model results from the peculiar cultural, political, legal and tax-related aspects of the current economic environment for the allocation, delivery and payment of health care in the United States. Underwriting individual group health coverage in an unregulated marketplace would be a relatively simple and standardized exercise. It is only due to the distorted structure of regulations and perverse incentives that dominate the health-care market that this model is necessary. As this structure of the health-care environment evolves and grows even more complex, the model will naturally have to adapt to match it.

(Finally, as a clarification, note that where the model refers to "the insurer" and "the insurer's gain," $G$ (in the positive sense), it actually represents the perspective of a self-insured employer. And the "gain," $G$, is actually the employer's "negative gain" or the portion of the plan's cost not funded by employee contributions. The employer's "negative gain" is not limited by its portion of the premium paid to an indemnity insurer but extends to the plan's individual benefit maximum, $M$.)

## (AUTHORS' REVIEW OF DISCUSSION)

CHARLES S. FUHRER AND ARNOLD F. SHAPIRO:

The authors thank Gerald Smedinghoff for his thoughtful discussion. His first refinement is very important, but it is outside the scope of the paper. Given today's cost of medical care, very few individuals would be willing to accept the risk of not insuring. Thus, practically all employees who decline medical insurance under a flexible benefit plan are covered under their spouse's plan. This explains our prediction of zero for the number who would choose reimbursement 4 in Section V of the paper. Of course, the methodology of the paper would extend to predicting choice with competing spousal plans. This probably should have been mentioned in Section VI.

The second refinement is a little less clear. First, the discussant seems to attach some meaning to the term: $A(i)$. In fact, he calls it "the cost deviation due to selection." A careful reading of our paper indicates that $A(i)$ is never so defined. As a matter of fact, $A(i)$ was defined for the purpose of illustration only. Perhaps the discussant intended to make the case that $A$, the cost deviation due to selection, should be defined differently. If this was his intent, he might have stated the exact definition. We defined $A$ so that it would satisfy some of the properties discussed in Section III.

Finally, we agree that we did sometimes use the word insurer when we mean the plan sponsor. Nevertheless, $G$, as defined, represents the excess of individual employee contributions over plan costs.

Once again, we thank the discussant for the enhancement and the clarification.