Research Poster Presented at the
American Psychological Society Annual Meeting
Washington, DC
May 1998

Pareto Group Instability and the Prediction of Health Care Claims Costs

Mark Attridge, Ph.D., M.A.

This study investigated the stability of heath care costs using a methodology featuring a two-year longitudinal design, stratified random sampling, a large sample size (N = 974), claims system data and self-report survey data, and statistical testing. The typical "regression to the mean" effect was observed, as extreme cases (both high cost and low cost) moved toward the middle during the next year. Almost two-thirds of cases changed their claims cost group status from one year to the next year. The "pareto" group (top 20% of costs in the past year) was the most unstable, with less than 4% still classified at the same highest-cost level the following year.   The most striking finding was that 92% of future claims costs could *not* be predicted, even when using past claims costs and relevant survey data on age, sex, health care visits and psycho-social concerns. Cost control implications of prevention, health promotion, and demand management services are discussed.

CONTACT INFORMATION:

Mark Attridge

Research Specialist / Project Manager

Optum® Division of United HealthCare Corporation

6300 Olson Memorial Highway

Mail Route MN010-S236

Golden Valley, MN 55427

(612) 797-2719

(612) 797-2730 FAX

www.optumcare.com

e-mail   mattridg@uhc.com

Pareto Group Instability and the

Prediction of Health Care Claims Costs

As the cost of health care has increased so have efforts to control costs. An assumption underlying many of these efforts is that health care costs can be predicted. A corollary assumption is that if costs can be predicted, then they can be managed.   One fact used to support this approach is that a small group of people typically do create the vast majority of health care costs. The 20/80 rule indicates that within a given population, 20% of the people create 80% of the costs for a year period, with this high-cost subpopulation being called the "pareto group." If one could predict who the pareto group will be, then the task of managing costs could be directed toward this smaller group rather than to all people in a health care system. However, the pragmatic utility of a pareto group is limited because by the time these cases have been identified they have already created large expenses.   Hence, the real trick is to find ways to predict who will become a high-cost pareto group member and then deliver successful interventions before these individuals fulfill their prediction for very high cost status.

The focus of this study is to explore the within-person stability of health care claims costs over time and to test if various relevant factors can predict claims costs in the next year.

Method

Survey Development

The survey used in this study was developed collaboratively by members representing three business units at United HealthCare (OPTUM Research, Total Care Management, and Analytic and Reporting Services) and the UHC of Oklahoma staff.

The need for an original survey was evident after a review of existing commercial health risk appraisal instruments found them unable to meet the multiple goals of the project. Several versions of the survey were written and refined before arriving at the final version, which was then converted to a machine-scannable format.

Sampling Procedure

Sample Size. The cost for surveying the entire population of all employees of the State (over 67,000) was prohibitively expensive and thus an alternative research design was chosen. This new design featured a stratified random sample of employees.   A sample size of around 1,000 people yields responses that have a low margin of item measurement error (plus or minus 3%) and past studies using a multi-stage survey process have found a return rate of about two-thirds. Thus, a target sample size of approximately 1500 employees was adopted.   It was expected that about 1,000 of these individuals would return the survey.

Sample Stratification by Claims Costs.   It is common for a small number of people to account for a large portion of the total costs for health care at a company. A purely random sampling process can yield few if any of the very high-cost cases.   To obtain a valid representation of the entire range of health care claim expenses among the sample who complete the survey, a claims cost-adjustment process was created.   This was done to stratify the sampling frame that determined who would be sent the survey. United HealthCare had the claims-paid information available for a period of time prior to conducting the survey study (i.e., May 1994 to June 1995).   Note that there was lag period of about 6 months before the survey was delivered that the claims costs were not available.

The amount of all paid claims for the year before were tabulated and for each employee (subscriber) the total costs for his or her family were divided by the number of insured family members. The entire population was then organized into five groups, ordered from the highest to the lowest amount paid as a percentage of all costs paid by the State (i.e., claims group 1 = the cases accounting for the top 20% of all costs; claims group 2 = cases accounting for the 61-80th percentile of costs; etc.). Note that less than one-half of one percent of families accounted for 20% of all claims and that only 8.7% of families accounted for 60% of all paid claims. Within each claims group, the goal was to randomly select an equal number of families.

TABLE 1. **Claims Groups Used For Stratified Sampling**

Claims Group	Percentile Range of All Claims Paid	Average Total Claims Costs Per Family Member Per Year	Cases as Percent of All Cases (67,318)	Number of Cases in Group	Number of Cases Selected for Survey
1	81-100	$52,669	0.4%	296	288
2	61-80	$13,550	2.1%	1,441	289
3	41-60	$4,668	6.2%	4,186	300
4	21-40	$527	56.1%	37,749	321
5	0-20	$0	35.1%	23,646	319
TOTAL		$1,107	100.0%	67,318	1,517

Survey Delivery.   A three-stage procedure was used for survey delivery.   The first phase involved mailing the survey (along with a cover letter and postage-paid reply envelope -- See Appendix) to the home of the insured employee.   At Phase 2, those who did not return the survey from Phase 1 were telephoned at home and asked to complete the survey over the phone.   At Phase 3, those who still had not responded by Phase 2,

were mailed a second copy of the survey and given a 12-page health information newsletter as a gift for their efforts.

Return Rate. A total of 974 surveys were returned (see Table 2). This is an overall survey return rate of 64%. Note that without the Phase 2 and 3 surveys, the overall return rate would have been just 33%, which is typical of past research using a one-time mail method of delivery.   Also, the return rate for a first contact home mailing was about the same as the more expensive method of calling people (Phase 1 = 33% vs. Phase 2 = 37%). The survey return rate at each phase did not differ by respondent age, sex, family size, or average claims costs.   However, the telephone method produced more spouses as respondents (13%) than did the two by-mail methods (both about 3% spouses).

TABLE 2. **Survey Return Rate**

	Phase 1. Mail	Phase 2. Phone	Phase 3. Mail again with incentive	TOTAL
Number Available in Sample	1517	1021	645	1517
Number Who Returned Survey	496	376	102	974
Percent of Total Returned	51%	39%	10%	100%
Return Rate as Percent of Available Sample	33%	37%	16%	
Return Rate as Percent of Total Sample	33%	25%	7%	64%

The sample who returned the survey did not differ from the rest of the total population (including those who were sent a survey but did not return it as well as those never sent a survey) in their average claims costs or family size (see Table 3). Thus, the survey sample is representative of the larger population for these characteristics.   Due to

the random selection process, it is assumed that the sample variability obtained for gender

and age is also similar to that of the entire population.

TABLE 3. **Comparison of Survey Participants and Non-Participants**

Claims Group	Average Total Claims Costs		Average Family Size (number of covered lives)	
	Returned Survey	Rest of Population	Returned Survey	Rest of Population
	(n = 974)	(n = 66,344)	(n = 974)	(n = 66,344)
1	$52,273	$53,495	1.21	1.25
2	$13,882	$13,499	1.41	1.39
3	$4,546	$4,674	1.53	1.53
4	$577	$526	1.77	1.89
5	$0	$0	1.55	1.56

The average age of the 974 respondents in the sample was 52 years (range 22 to

92).   Sixty-five percent were female and 35% were male.   Of respondents, 665 (68%)

had a spouse and 332 (34%) had one or more children.

Measures

The survey instrument was designed by the author for United HealthCare

Corporation (see Appendix). It asked if the person had any of 14 common health

conditions (heart disease, cancer, asthma, etc.), how often the person visited any kind of

health care provider in the past year, and ratings of concern in 10 areas of life (physical

well-being, work, money, family, personal relationship, etc.) associated with use of

employee assistance services.

Results

Research Question 1:   How stable are claims costs from one year to the next year?

The change in claim cost group status from one year to the next was examined

first (see Table 4).   Group status in 1996 was defined using the same dollar ranges used

to create the five groups in 1995. Using this standard, only 36% of families remained in

the same cost group from one year to the next; 59% decreased; and 5% increased.   Of the

200 families in the pareto group in 1995, only 7 were at that same cost level in 1996.

Thus, just 3% of cases at the pareto group highest-cost status remained at that high level

for the next year. Clearly, identification of the highest cost cases has little to no utility in

determining who will be the highest cost cases the next year.

TABLE 4   **Change in Claims Costs Group Status 1995 to 1996**

		1996 Group					
Range of annual per person total paid claims	1995 Group	1	2	3	4	5	total n
$0	1	**147**	24	0	2	0	173
$1 to $2,900	2	30	**172**	7	5	1	215
$2,901 to $8,000	3	18	157	**12**	6	1	194
$8,001 to $28,000	4	20	142	16	**13**	1	192
$28,027 to $217,907	5 Pareto	23	115	31	24	**7**	200
	total n	238	610	66	50	10	974

Note: 1995 Group dollar ranges used for assignment to 1996 Group. On diagonal = no change
from 1995 to 1996 (36%); below diagonal = less costly in 1996 than 1995 (59%); and above
diagonal = more costly in 1996 than in 1995 (5%). All 974 cases continuously enrolled for the
two years of the study.

The over time instability in health care claims costs was also reflected in a

statistically significant but rather low correlation between 1995 costs and 1996 costs of

$r = .25$.   This correlation indicates that knowing last years costs accounts for just 6% of

the total variance in next years costs.   Paired t-test analyses conducted within each 1995

cost quintile group (see Table 5) indicated that while the two low cost groups had

nonsignificant small increases in average costs, the three higher cost groups had very

large decreases in average costs (i.e., from four to ten times lower).

TABLE 5   **Average Annual Per Member Claims Costs 1995 and 1996 by Group**

Range of annual per person total paid claims in 1995	1995 Group	1995 Costs Mean (*SD*)	1996 Costs Mean (*SD*)	*t* value difference Year to Year	*r* correlation Year to Year	*n* *sample size*
$0	1	**$0** ($0)	**$156** ($1,224)	-1.68 ns	N/A	173
$1 to $2,900	2	**$577** ($709)	**$972** ($3,337)	-1.77 ns	+.22***	215
$2,901 to $8,000	3	**$4,546** ($1,332)	**$1,256** ($3,258)	13.10***	+.02 ns	194
$8,001 to $28,000	4	**$13,882** ($5,336)	**$1,868** ($4,165)	27.57***	+.21***	192
$28,027 to $217,907	5 Pareto	**$52,273** ($33,093)	**$4,780** ($12,156)	19.53***	+.08 ns	200
Total	All	**$14,503** ($24,67)	**$1,842** ($6,400)	-16.30***	+.24***	974

Note: All 974 cases continuously enrolled for two years of the study.
ns = not significant at $p < .05$.   *** $p < .001$

Research Question 2:   How much of next year's claims costs can be predicted?

A multiple regression analysis was used to predict 1996 average per family member claims costs. The predictor variables included 1995 measures of total per family member paid claims costs, respondent age, respondent sex, respondent number of health care visits, respondent number of health problems and conditions, and respondent average ratings of 10 areas of life concerns (see Table 6).

TABLE 6   **Zero-order Correlation and Multiple Regression Results for Predicting Future Health Care Claims Costs**

						Dependent Measure 1996 Claims Costs	
		correlations					
1995 Predictor Measures	1	2	3	4	5	r	Beta
1. 1995 Claims Costs						.25*	.22*
2. Age	.16*					.01	-.03
3. Sex (1=male 2=female)	.02	-.16*				.01	-.01
4. Visits to Health Care Provider	.26*	.05	.16*			.18*	.12*
5. Number of Health Conditions	.15*	.25*	.01	.33*		.09*	.02
6. Life Concerns	.01	-.26*	.06	.08	.10*	.02	.00
Multiple Regression R							.28*
R squared							.08*

Note: N = 894. Variables 2 to 6 are all self-report data.   * p < .05

Results revealed that this model was significantly able to predict next years costs, however only 8% of the variance in future costs was accounted for. Thus, 92% of the variance in next years health care costs were <u>not</u> able to be predicted by this set of predictor measures. Future costs were predicted best by using the measure of last years costs (standardized regression coefficient Beta = .22).   Self reported total number of

visits to health care providers in the past year was also predictive of next years claims

costs (Beta = .12). The lack of predictive value of perceived life concerns is surprising,

especially when psycho-social factors are increasingly found to play a critical role in

overall health.   Demographic factors of age and sex also did not predict claims costs.

Conclusions

This longitudinal study used a well-developed methodology featuring stratified

random sampling, a large sample size, claims system data and self-report survey data, and

statistical testing to investigate the stability of heath care costs.   Almost two-thirds of

cases changed their claims cost group status from one year to the next year.   The typical

"regression to the mean" effect was observed, as extreme cases (both high cost and low

cost) moved toward the middle of the cost distribution during the next year. The "pareto"

group (top 20% of costs in the past year) was the most unstable, with less than 4% still

classified at the same highest-cost level the following year.   The most striking finding

was that 92% of future claims costs could *not* be predicted, even when using past claims

data and relevant survey information.

In general, the evidence from this study suggests that attempts to predict, much

less control and manage, future health care costs will be largely unsuccessful due to the

basic instability of cost generating behavior at the individual level.   This is a sobering

finding for managed care companies interested in controlling health care costs in the

short-run of a single year. Note that short-term cost control pressures are real as most

health care insurance business partnerships are on a one year contract.

One implication is that trying to use claims and survey methods to identify, target,

and intervene on a small "hi-risk" segment of the population may fail precisely because

the people identified are unlikely to stay at that high-cost level and furthermore, such targeting may actually exclude the very people who will become high cost cases in the near future. An alternative strategy for attempting to control of future health care costs may be to provide health promotion and utilization demand management services to the entire group membership (rather than just a targeted "hi-risk" pareto subgroup). The "cast a wide net" approach is based on a prevention model that is also simpler (and often cheaper) than having to institute and coordinate various procedures for claims analysis and/or survey methods of identifying and targeting small groups of a population for cost-control interventions.

Given the instability and the complexity in health care seeking tendencies of individuals, it may be wise to invest in benefits services that are offered to everyone, available whenever they are needed (24 hour 7-day a week access) and that provide information and appropriate referral resources for health care services.   Although such services are typically not used by everyone in a population, the 10% to 30% of members who do use them are self-selected (each person decides to use the service and initiates contact) due to an emergent and, this is key, *unpredictable* health care need.   The merit of this approach is starting to be validated in recent empirical research showing overall cost savings and positive return on investments for these kinds of services (Bergmark et al., 1996; MacStravic & Montrose, 1998; Otis et al., 1997; Parker, Bergmark, & Attridge, 1997).

References

Bergmark, R. E., Dell, P., Attridge, M., Parker, M. (1996). Creating an integrated health care system: The health and human risk model. *Managed Care Quarterly*; 4(1): 36-42.

MacStravic, D. & Montrose, G. (1998). Managing health care demand. Gaithersburg, MD: Aspen.

Otis, J., Kelley, B., Jacobs, A., Attridge, M. (1997). "Two-year effect of a demand-side management program on outpatient utilization--a summary of findings to date." In *Faulkner & Gray's Guide to Managed Care Strategies 1998*.   Faulkner & Gray, NY.

Parker, M., Bergmark, R. E., & Attridge, M. (1997). "The mind-body connection: Outcomes research in the real world." In J.D. Haber & G. E. Mitchell (Eds.), Primary care meets mental health: Tools for the 21st century. CentraLink Publications: Tiburon, CA.