

Analyzing Healthcare Costs with SAS: An Intern's Experience

Ben Keefer, The Regence Group, Portland, OR

Abstract

The goal of this analysis was to measure the effect of Regence's wellness program on healthcare claims costs. This work will serve as a framework to establish formal statistical models of Return on Investment (ROI) and will indicate which activities yield the highest positive return. ROI is defined here as the ratio of savings in claims costs from participation to the program's costs in dollars (savings / costs). A positive ROI simply means that every dollar spent on the wellness programs generates more than a dollar's worth of savings.

Complications and Insights

While the ROI equation is simple and straightforward, the definition of these metrics is complicated. To understand "savings" associated with the wellness program, we must explore how healthcare costs differ depending on participation.

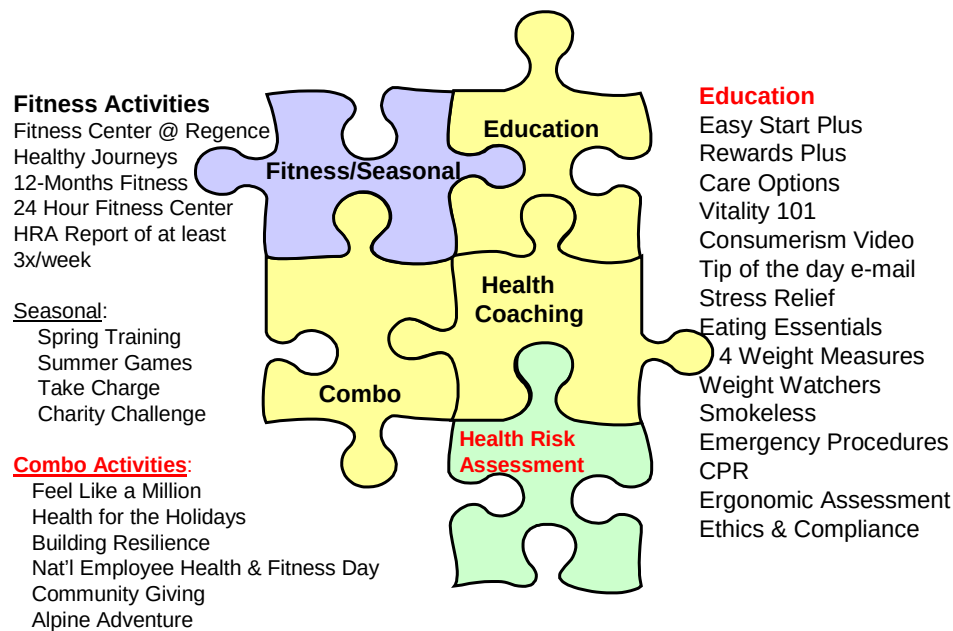
Two questions quickly become apparent. First, when does the impact of the intervention manifest? Second, is it possible that in the short-term costs might increase due to positive utilization or preventive care while long-term costs decline? Because the timing of the effect is unknown, it would be helpful to chart the costs before and after the intervention. Also, because of the possibility of positive utilization—which would show up under the outpatient and professional services components of healthcare costs—it would be helpful to see how the components are also affected. Corporate reporting generally requires concise display of the data; therefore, the challenge is to convey the greatest amount of information in the fewest charts possible.

Understanding the Data

The most important components of the data are the interventions captured and the cost data available.

When trying to make sense of the interventions offered, the first difficulty is the sheer number of interventions. With over thirty interventions spread out over three years, there are too many to test individually, making it necessary to group similar activities together.

One way to group similar activities is by type, as shown in the chart below.



In the chart, the puzzle pieces show five different groups of interventions, with the underlying components listed off to the side. The Health Risk Assessment and Health Coaching are the only two groups for which there are no underlying activities. Wellness programs were originally designed to target cardiac risk factors. Therefore, each group or bucket was defined with a separate hypothesis regarding the effect participation would have on total claims costs. For example, it was hypothesized that fitness related activities would have a different impact than educational activities. Rather than testing the components individually, we can now test for the impacts for different groups of interventions.

Now that the activities are grouped into different intervention buckets, how should we try to capture the nature of the intervention? Previous studies have used intensity or strenuousness of the activity, with fitness more intense than educational activities. Others have instead tested the duration or number of years of participation. Finally, we could also measure the frequency of participation within each wellness bucket.

One compromise would be to separate individuals based on the total points accumulated for the different activity groups. At Regence, points are accumulated for completion of activities—with each activity having a different number of points assigned to it—and gift cards are awarded at specific point levels. Total points captures information of frequency and duration simultaneously. For example, if an individual participated in multiple activities in the same year or the same activity for multiple years, he would have accumulated more points than someone who participated in a single activity for a single year.

Having dealt with the intervention, the next question is how to measure the impact. We have two types of variables available: costs for the year or cost per month. Although costs per months control for differences in the months enrolled in a Regence's health insurance plan, they are harder for a general audience to understand. Consequently, costs for the year will be used. As discussed previously, we will be interested in the following components of total yearly costs: inpatient, outpatient, professional services, and pharmaceutical.

SAS Code

Since the data is structured with each observation representing a unique individual and different variables capturing cost and intervention data, simple SAS procedures will help get the information organized.

Quintiles are often used as a method of analyzing data where solid endpoints of the effect on an outcome are unknown. Proc Univariate and Proc Print were used to get the data needed to create quintiles on total education points, while Proc Means was used to display the final results. The code used to measure the intervention on the education bucket is given below.

```
proc univariate data = wproisets3b;
    output out=distrib pctlpts=20 40 60 80 pctlpre=educ;
    var educ ;
run;

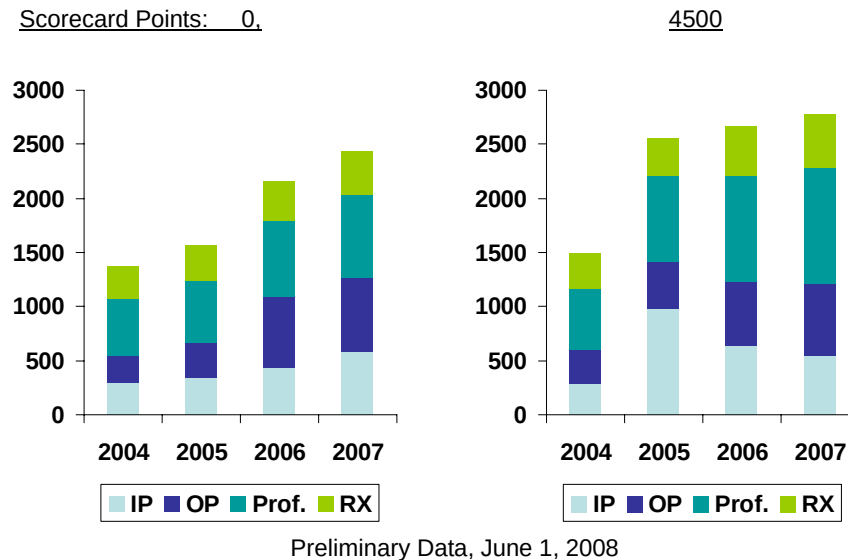
proc print data=distrib;
    title 'Quintiles for Educ Points';
run;

data wproisets3b;
    set wproisets3b;
    if educ=0 then QuintileEduc=1;
    if 4500<= educ then QuintileEduc=5;
run;

proc means data=wproisets3b;
    class QuintileEduc;
    var inadpay04 inadpay05 inadpay06 inadpay07
        inadipay04 inadipay05 inadipay06 inadipay07
        inadoppay04 inadoppay05 inadoppay06 inadoppay07
        inadphpay04 inadphpay05 inadphpay06 inadphpay07
        inadrxpay04 inadrxpay05 inadrxpay06 inadrxpay07;
run;
```

After executing the code, Excel was used to create the following chart.

Cumulative Costs and Utilization Shift over Time:



The chart shows that for people without any participation in educational programs, each inflation-adjusted cost measure increased in a roughly linear fashion for the years in our study. In contrast, those with more than 4500 educational points (the top quintile) saw inpatient costs decline, with outpatient and professional costs representing a larger portion of total costs. This suggests that these individuals are likely using preventive care, which should result in long-term cost savings.

Conclusion

Although the task of analyzing the effect of healthcare interventions may seem daunting, simple SAS procedures like Proc Univariate and Proc Means are powerful tools to help develop a framework for further statistical modeling. In the case of Regence's wellness program, these procedures have provided compelling evidence that educational activities significantly impact healthcare costs.

References

Naydeck BL, Pearson JA, Ozminkowski RJ, Day BT, Goetzel RZ. The impact of the Highmark employee wellness programs on 4-year healthcare costs. *J Occup Environ Med.* 2008;50:146--156.

Acknowledgments

The author is deeply grateful to the support of The Regence Group and supervisor Barbara Naydeck for providing the opportunity to conduct this research. Also, the author

would like to recognize the contributions and guidance of fellow researchers Kate Turyn and Hong Chen.

Contact Information

Your comments and questions are valued and encouraged. Contact the author:

Ben Keefer
The Regence Group
100 Southwest Market Street
PO Box 1071 | M/S E10G
Portland, Oregon 97207-1071
503-525-6577
bpkeefe@regence.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and all other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.