

**Title:** An Ounce of Prevention Worth a Pound of Medical Care? Medical Utilization Under A Wellness Plan

**Author:**

Shooshan Danagoulian  
Department of Economics  
Wayne State University

**Corresponding Author:**

Shooshan Danagoulian  
Department of Economics  
Wayne State University  
656 W. Kirby St.  
Detroit, MI 48202  
Email: fr4523@wayne.edu  
Tel: 313-577-2693

**Keywords:** Preventive Care, Medical Utilization, Health Insurance

**JEL Codes:** I130, I120, G220, I110

**Abstract:** The increased prevalence of employer sponsored wellness programs is intended to steer employees towards healthier lifestyle and improved maintenance of chronic conditions. This paper studies the medical expenditure and utilization of a health insurance plan with wellness features offered by a large self-insured employer. The analysis uses difference-in-differences estimation of utilization by wellness members matched to non-members using propensity score matching. The results show that while the wellness program increases utilization of preventive and outpatient care, by as much as 1.5 visits per year, there is no comparable decline in emergency or inpatient care, resulting in an overall increase in medical expenditure of around \$680. The increase in medical expenditure persists even 6 to 7 years of continued enrollment in wellness, suggesting that the hypothesized future reductions in medical expenditure are not likely to materialize.

# 1 Introduction

More than 50% of US employers offer workplace wellness programs in response to the rising incidence of chronic disease due to lifestyle conditions such as inactivity, poor nutrition, tobacco use, and frequent alcohol consumption. A workplace wellness program can include discounted access to gym, health and lifestyle counseling, as well as partnerships with local area physicians to increase the use of preventive care services. Such programs are intended to provide resources for healthy living, to identify and treat health risks, and to improve the maintenance of chronic conditions. They may be provided as a stand-alone benefit or as an optional or mandatory component of employer provided health insurance. The Affordable Care Act of 2010 (ACA) provides tax benefits for adoption of wellness programs, and allows employers to charge employees a higher premium on insurance for tobacco use, as long as such a program is offered with smoking cessation support.

The benefit of a wellness program to employers is two-fold. The employer would save through improved health of its employees and the directed health services utilization. Improved health and chronic conditions maintenance would reduce the incidence of absenteeism and presenteeism, creating productivity gains for the employer. The identification and improved maintenance of health risks would shift the medical care away from high cost inpatient and emergency room care towards lower cost outpatient and preventive care. Both changes should translate into reduced health care costs for the employer.

This paper evaluates the use of healthcare services in a health insurance plan with an integrated wellness program. The data used in the analysis are health insurance choice and utilization claims from a large self-insured university employer. The employer introduced a health insurance plan with an integrated wellness program in 2008, Aetna Wellness. The plan was introduced in parallel to the same plan without the wellness features, Aetna. Aetna Wellness required the choice of a primary care physician and the completion of a health risk assessment questionnaire. Despite sub-

stantial financial incentives, such as lower premium and deductible, enrollment in Aetna Wellness grew slowly.

Our results show that enrollment in the Aetna Wellness does not result in significant decline in overall expenditure, but does lead to up to \$639 increase in outpatient care per person per year for members who have been enrolled for 4 years or more. In terms of visits, members have 0.45 more preventive visits per year, part of a 1.57 visit increase in their overall outpatient visits. While Aetna Wellness members are 3 percentage points less likely to use the emergency room, they are 13 percentage points more likely to seek out preventive care. Over a larger period of time, the increased preventive care use persists without a comparable decline in acute or emergency care. While women make up a larger share of Aetna Wellness members, the rise in utilization is not driven primarily by women.

In economic literature, wellness programs are a type of prevention viewed through the Grossman (1972) human capital model. In this model, investment in health increases utility and time available for productive activity. However, investment in health can be preventive and curative, which are substitutes as described by Grossman and Rand (1974). Health insurance presents an *ex ante* moral hazard by increasing the cost of preventive care relative to curative care, resulting in underutilization of preventive care (Pauly 1986, Ehrlich and Becker 1972, Breyer and Zweifel 1997). Reducing the monetary and time costs of preventive care makes the substitution from curative care easier (Shavell 1979, Tolley, Kenkel and Fabian 1994); a market solution to the moral hazard problem would be a risk-rated health insurance, such as one where smokers pay a higher premium than non-smokers (Kenkel 2000).

Few studies evaluate workplace wellness programs in economics literature. Among these, Mukhopadhyay and Wendel (2013) evaluate an employer wellness plan three years after its implementation. They find no change in healthcare utilization or absenteeism, but do see increased

rates of recommended health screenings. In the specific outcome of weight loss, Cawley and Price (2013) find that employer wellness plans with financial incentives have very high attrition rates and result in modest short term weight loss. Downey (2014) summarizes this literature by noting that, particularly in weight loss, employer wellness programs have not been able to generate long term changes.

Oscilla et al. (2012) conducts a systematic review of efficacy studies of wellness programs. Among the studies which looked at healthcare costs and utilization different methods of evaluation were used: randomized control trial (Milani and Lavie 2009), nonrandom comparison groups using propensity score matching (Henke et al. 2011, Merrill et al. 2011, Naydeck et al. 2008), and descriptive studies (Aldana et al. 2005, Stave et al. 2003, Yen et al. 2010).

The studies closest to the current paper are Henke et al. (2011) and Naydeck et al. (2008). Henke et al. (2011) compared the wellness program at Johnson & Johnson to sixteen other firms in comparable industries, some of which had wellness programs of their own. The authors looked at medical costs between 2002-2008, in the third decade after the introduction of the wellness program, for employees enrolled at least 2 years in the program. They find 3.7% lower average annual growth in medical costs, and annual savings in medical costs of about \$535 per employee.

Naydeck et al. (2008) conduct a similar study for Highmark Inc. employees over a 4 year period following the introduction of a wellness program. They find \$176 per person lower medical expenses for participants, and \$182 per person per year savings in inpatient expenses.

This paper contributes to the literature in three ways. First, it tracks medical utilization 4 to 7 years after enrollment in a wellness plan, the longest panel analysis to our knowledge. This is a particularly important in evaluation of wellness programs as preventive care measures are not likely to show results in the immediate horizon. Second, it analyzes the composition of medical

services utilization in expenditure and actual visits, in addition to total expenditure. Third, this paper combines propensity score matching with difference-in-differences estimation method to control for the selection bias present in the choice of health plan and medical utilization. Thus, our analysis allows us to compare a substantial period of pre-wellness utilization to post-wellness utilization.

The remainder of this paper will proceed as follows: Section 2 will introduce a model of preventive care. Section 3 will discuss the estimation challenges the enrollment in Aetna Wellness poses for the analysis. It will also discuss the method used to alleviate the selection occurring in the estimation. Section 4 will discuss data and present summary statistics. Section 5 will discuss the estimation strategy. Section 6 presents the results and Section 7 concludes.

## 2 Model

The Grossman (1972) health capital model gives structure to modeling demand for preventive care. The individual derives utility from a stock of health,  $H_t$ , at time  $t$ , and a composite good,  $Z_t$ .

$$U(H_t; Z_t) \quad t = 0, 1, \dots, n$$

The individual is born with an initial endowment of health,  $H_0$ , which depreciates at a rate  $\delta$ . The individual can invest in their health to offset the depreciation, creating a law of motion for the stock of health.

$$H_{t+1} = I_t(M_t, Pv_t, TH_t; \epsilon_t) + (1 - \delta)H_t$$

where  $M_t$  is medical care,  $Pv_t$  is preventive care,  $TH_t$  is time spent on health generating activities, and  $\epsilon_t$  is an iid shock. Aside from delivering utility, the stock of health reduces the ill time the

individual experiences, thus increasing available time for work and, hence, consumption of the composite good. Given a total hours of time,  $\Omega$ , the time constraint is defined by

$$TW_t + TH_t + TL_t = \Omega$$

The individual maximizes utility subject to a budget constraint

$$\sum_{t=1}^n \frac{P_t M_t + V_t P V_t + Q_t Z_t}{(1+r)^t} = \sum_{t=1}^n \frac{W_t T W_t}{(1+r)^t} + A_0$$

where  $P_t$ ,  $V_t$ ,  $Q_t$  are the respective prices of medical care, preventive care, and composite good,  $W_t$  is the wage,  $r$  the economy-wide interest rate, and  $A_0$  is the initial asset endowment.

The ill time is a function of health stock

$$TL_t = g(H_t)$$

where  $g()$  can be a linear function as in the Grossman model, or a negative exponential function. In both cases, however, ill time and health stock are negatively related, so that  $\partial TL_t / \partial H_t < 0$ .

The lifetime utility is maximized with respect to  $M_t$ ,  $P V_t$ ,  $Z_t$ ,  $TH_t$ , and  $TL_t$ . The first order conditions can be reduced to

$$\frac{\pi_t}{(1+r)^t} = \sum_{i=1}^{n-t} (1-\delta)^{i-1} \left[ \frac{U h_{t+i}}{\lambda} + \frac{g' W_{t+i}}{(1+r)^{t+i}} \right] \quad (1)$$

Here, the present value of  $\pi_t$ , the marginal cost of gross investment in health in period  $t$ , is equal to the present value of benefits. The first term in the brackets is the discounted monetary value of the increased utility due to the increase in health stock, while the second term is the discounted monetary value of the increased work time due to increased health stock.

Equation (1) implies that as the price of preventive care declines,  $V_t$ , quantity demanded will increase. Declining cost of preventive care allows the present discounted marginal benefit of care to decline, and, thus, the individual is willing to engage in more preventive care.

The model predicts that the introduction of Aetna Wellness, as it reduces the monetary and time costs of acquiring preventive care some prefer will increase demand for it. This model does not allow an evaluation of the relationship between medical care and preventive care. Thus, assuming that the cost of medical care does not change, the model predicts no change in demand for it. As a result, there should be an overall increase in use and expenditure of all health care services, albeit with some substitution between preventive and medical care.

The next section discusses the Aetna Wellness plan and the estimation challenges the choice of insurance presents.

### **3 Aetna Wellness**

In 2008, in cooperation with the local area physicians, the employer introduced Aetna Wellness to the health insurance menu on offer to its employees. The aim of the Aetna Wellness plan was to include the primary physician more actively in the preventive health care for the member, to identify and treat health risks before they evolve into costly hospitalizations. The plan was also designed to encourage healthier lifestyle by offering discounts to the gym, counseling for obesity, smoking cessation, and stress management. To achieve these goals, the enrollees were required to choose a primary care physician (PCP) and complete a health risk assessment questionnaire. On the basis of the questionnaire, the PCP would conduct an annual physical examination. The PCP and the member would develop a health plan, which might include counseling sessions such as smoking cessation, nutrition, stress management at a lower copay, ranging from \$10 to \$20,



compared to the other plans. Beyond the initial examination, the PCP did not act as a gatekeeper for the health plan as the member could use any specialist services without referrals, and Aetna Wellness was a preferred provider organization (PPO) plan.

Aside from Aetna Wellness, employees could choose from an identical Aetna plan without wellness features, and another PPO plan with a different network of physicians and hospitals. Given the geographically concentrated nature of the employee residence, a comparison of the in-network doctors the most frequent zip code of residence shows that the networks were sufficiently similar not to have a significant effect on utilization of medical services. All three of the plans had the same coinsurance rate, copay, and pharmacy benefits. They differed in the premium and the deductible. See the Appendix 2 for a more detailed comparison of the plans, and a discussion of other plans offered by the employer.

Because enrollment in Aetna Wellness was voluntary, its members may have selected themselves into the plan based on expectations of future utilization or salience towards the features of the plan. More health conscious employees may choose to enroll in Aetna Wellness to take advantage of gym discounts and cheaper preventive care. This would place downward bias on estimates of change in medical utilization as members are healthier than the average employee. On the other hand, Aetna Wellness may attract employees with more chronic conditions, looking for cheaper and better maintenance of those conditions, resulting in an upwards bias of estimates. Danagoulian (2016) explores the selection into a wellness program, and finds some effect of health on probability of enrolling into the plan.

To address the selection bias, two techniques are combined to account for the observable and unobservable factors of selection. Aetna Wellness members are matched to non-members using propensity score techniques developed by Imbens and Rubin (2015), accounting for the observable heterogeneity between the treated and the control. The change in the utilization is then estimated

using difference-in-differences estimation of the matched samples, removing the unobservable heterogeneity. The following section discusses matching and difference-in-differences techniques to mitigate selection.

## 4 Methods

In an ideal experiment, employees would be assigned randomly to the Aetna Wellness or another plan, and their medical utilization would be compared. Random assignment ensures conditional independence of the outcome from enrollment in the treatment. Let  $Y_{i1}$  represent the utilization when employee  $i$  is enrolled in the wellness plan (treatment), and let  $Y_{i0}$  be the utilization of individual  $i$  in the other plans (control). In a random assignment setting, the comparison of outcomes gives:

$$\tau = E(Y_{i1}|T_i = 1) - E(Y_{i0}|T_i = 0) = E(Y_i|T_i = 1) - E(Y_i|T_i = 0)$$

where  $\tau$  is the treatment effect and  $T_i$  is the treatment status. That is, if the treatment is randomly assigned, then the difference in the observed utilization between the treatment and control groups is a consistent estimate of the treatment effect

In observational data, such as those used here, the treatment and control groups are not randomly assigned. The failure of the conditional independence assumption implies that the difference in utilization between the treated and the control cannot be attributed to the treatment alone. In particular, without conditional independence, the factors which determine medical utilization may also determine the choice to enroll in the Wellness program. This means that the treatment group is substantially different from the control group, and the difference, in part, determines the observed outcome.

Rosenbaum and Rubin (1983) proposed that conditional on the observable characteristics of individual  $i$ ,  $X_i$ , there should not be a systematic pre-treatment difference between the groups assigned to treatment and control. To overcome the dimensionality of the characteristics, they propose the propensity score as the means by which to match the treatment with a control group. Let  $p(X_i)$  be the probability of individual  $i$  enrolling in treatment, defined as  $p(X_i) = Pr(T_i = 1|X_i) = E(T_i|X_i)$ , and  $0 < p(X_i) < 1$ . If  $\{(Y_{i1}, Y_{i0}) \perp\!\!\!\perp T_i\}|X_i$ , then the treatment effect,  $\tau$ , can be defined as:

$$\tau|_{T=1} = E\{E(Y_i|T_i = 1, p(X_i)) - E(Y_i|T_i = 0, p(X_i))|T_i = 1\}$$

where the outer expectation is over the distribution of  $p(X_i)|T_i = 1$ . That is, conditional on the propensity score as function of observable characteristics, the treatment effect is the difference in observed outcomes between the treated and control.

Rosenbaum and Rubin (1984) addressed the choice of covariates  $X_i$  in the estimation of the propensity score, and proposed a stratification method adding higher order terms and interactions until there is no significant difference between the two groups. Imbens and Rubin (2015) and Imbens (2015) formalized balancing and stratification of the control group to improve matching on propensity score or covariates. The analysis here will follow the steps outlined by Imbens and Rubin (2015).

One of the main criticisms of propensity matching is that it matches on the basis of observable characteristics only. The quality of the match depends on the extent of matching characteristics available and on the sample size. At its best, however, propensity matching cannot overcome the selection into treatment based on unobservable characteristics. To mitigate this Heckman, Ichimura, and Todd (1997, 1998) propose combining a difference-in-differences estimator to propensity matching. Smith and Todd (2005) show that while propensity score matching may not be the best estimator in general, difference-in-differences matching estimators performed sub-

stantially better than cross-sectional matching estimators. Furthermore, Smith and Todd find little difference in estimation bias between various matching procedures.

Since the data used here track employees at a single university across years, the panel structure of the data combined with the institutional framework justifies the use of difference-in-difference estimator with propensity score matching. The next section will introduced the specifications to be estimated.

## 5 Estimation Strategy

The estimation of the propensity score will follow steps outlined by Imbens and Rubin (2015). In the first step, the matching covariates are selected in an iterative procedure, for the linear, quadratic, and interactive terms. In the second step, the sample is trimmed in the overlap region dropping extreme values from the control group. The final step consists of block balancing the treated and control groups using an iterative procedure. To compare the matched observations in regression analysis weights are generated using kernel matching of propensity scores. A detailed analysis of the matching process is presented in the Appendix 3.

Before analyzing utilization in a difference-in-differences specification, an event study will allow to visualize the trends in expenditure among those who switched into Aetna Wellness sometime during panel window, and those who did not. An event study estimates the difference in expenditure between treated and control in the time periods leading up to, and following the onset of treatment. To do so, the following model is estimated:

$$Y_{it} = \alpha + \sum_{j=T_{min}}^{T_{max}} \beta_{1j} T_i \cdot Time_{ij} + \beta_2 Health_{it-1} + \gamma X_{it} + \epsilon_{2,it} \quad (2)$$

Here,  $T_i$  is the indicator for those who switched to Aetna Wellness at any time during the panel period, and  $Time_{ij}$  is a dummy variable for each year of data before and after switching. The timing of this variable is important, as it tracks the years before and after switching for each individual:  $T_{min}$  corresponds to the largest number of panel years prior to switching, and  $T_{max}$  is the largest number of years after switching. The resulting vector of  $\beta_1$  estimates the difference in outcome variable between the treated and control in each year prior and post switching.

The difference-in-differences specification will compare Aetna Wellness members' utilization before and after switching in relation to the utilization of non-members over the same period of time. The specification to be estimated is:

$$Y_{it} = \alpha + \beta_1 AetnaWellness_i + \beta_2 AetnaWellness\_After_{it} + \sum_{j=1}^7 \beta_{3j} Year_{jit} + \gamma X_{it} + \epsilon_{it} \quad (3)$$

where  $AetnaWellness_i$  is an indicator if the individual ever switched into Aetna Wellness,  $AetnaWellness\_After_{it}$  is equal to one if the individual is currently enrolled in Aetna Wellness, and  $Year_{jit}$  is a year fixed effect. Since the panel data are not balanced, and the switch to Aetna Wellness is not at the same time for all individuals, the fixed year effects capture the change in characteristics common to the treated and the control group.

## 6 Data

The data consist of health insurance plan selection and claims data from a large self-insured university, spanning the period of 2007 to 2014. The sample consists of the universe of eligible employees and their families who have been enrolled in Aetna Wellness for 4 years or longer. The employees include active employees both hourly and salaried, administrative and faculty. The sample consists of 35,428 person years, following 7,630 employees and families for an approximately 4.6 years.

Among these 1,165 individuals switch to Aetna Wellness and have been enrolled for 4 years or more, and are followed for an average 7.1 years, for a total of 8,276 person years.

Since enrollment was on a rolling annual basis starting in 2008, the treatment group is defined as any individual enrolled in Aetna Wellness between 2008 and 2014. The control group, therefore, consists of any individual who was never enrolled during the panel time. The unbalanced panel structure also precludes a before and after classification for the control group, necessitating the year fixed effects in the specification.

---

Table 1 about here

---

Table 1 presents the summary statistics for the treated and control groups. The treated and control are statistically significantly different in most dimensions. The Aetna Wellness enrollees tend to be slightly older, predominantly female, with more dependents. In this sample, Aetna Wellness employees also earn a higher salary.

To control for the health status of individual, two indices are used. The Charlson index of comorbidities assesses the presence of more severe medical history such as cerebrovascular disease, congestive heart failure, and diabetes; as such it does not have sensitivity in the majority of the sample, but captures the health status of the highest spending group. The index ranges from 1 (least severe) to 6 (most severe), with a score of 0 if no conditions are present. Since more than 87% of the sample has a score of 0, the average score between the treated and control is fairly small, though statistically different.

The Medical Utilization Index is generated by a case-mix software developed at The Johns Hopkins Bloomberg School of Public Health. The Johns Hopkins ACG<sup>®</sup> System (Version 10.0.1) uses diagnostic and pharmacy codes from claims data to generate a five scale medical resource utilization index. The ACG System provides an output which categorizes the individual according to current health services utilization on a scale from 0 to 5, with 0 score indicating no diagnosis available, and 5 indicating a very high user. With more than 60% of the sample having a score of 0, the average medical utilization intensity is greater among the treated than the control.

While the total medical expenses are not significantly different between the groups, the Aetna Wellness members have a statistically lower emergency, acute, and pharmacy expenses. They have somewhat higher other expenses which include outpatient and preventive care.

A similar pattern emerges when looking at Aetna Wellness members before and after switching. Table 2 summarizes the medical expenditure by category before and after switching. Total medical expenditure increases by average \$712 per individual, of which other medical expenses account for \$613.

---

Table 2 about here

---

The summary statistics suggest that while there is a difference between the treated and control groups, the overall impact of the Aetna Wellness plan appears to be rising outpatient and preventive expenditure, without the compensating decline in inpatient and emergency care in the 4 to 6 year period following enrollment.

The outcome variables of interest focus both on expenditure and on utilization. The data provide several categories by type of medical expenditure: total, emergency, pharmacy, and inpatient acute. The category of other expenditure is defined as the difference between the total and all other categories. All expenditure variables are defined as the insurance negotiated price for procedure, and include both the patient and insurance share of the cost. Utilization is also analyzed as visits by category on both the intensive as well as extensive margin: emergency, preventive, inpatient, outpatient, nutrition counseling, and mammogram. To generate the visits information, the claims data are catalogued by type. To define the emergency, inpatient, and outpatient visits, the place of service is used as the indicator of visit type. Thus, the place of service for emergency visit is a hospital emergency room, while for inpatient it is an inpatient hospital. An outpatient visit can take place in a number of locations, including outpatient hospital, ambulatory surgical center, or doctor's office.

To define the preventive service category, an Aetna Benefit Guidance Statement was used, which listed all ICD-9 diagnostic codes which are covered as preventive. Aside from wellness exams, these include developmental testing; screening for depression, hearing, vision, cancer, osteoporosis, anemia, sexually transmitted diseases, cholesterol, diabetes, sickle cell, hepatitis B and C, HIV, and many others; immunizations; electrocardiograms; obesity preventive counseling; healthy diet counseling; alcohol/drug counseling; tobacco counseling; sexually transmitted infections counseling; and genetic counseling.

## **7 Results**

The event study allows us to track the changes in medical expenditure before and after switching to Aetna Wellness. Figure 1 presents the event studies for four categories of medical expenditure:



total, emergency, pharmacy, and inpatient acute care. Each panel depicts the estimated value of  $\beta_{1j}$  in equation (2), that is, each point estimate depicts the difference in expenditure between the eventual Aetna Wellness member and those who never switched into the plan. Each specification controls for individual health, insurance plan deductible for in-network care, gender, whether the person is a dependent, the number of dependents in the family, whether the employee is faculty or administrative staff, and race. In each specification, the treated observations are matched to control observations. While the sample is restricted to individuals who have been enrolled in Aetna Wellness for more than 4 years, the estimates track utilization in the years leading up to and following the switch, resulting in estimates for up to 4 years leading up to the switch, and up to 7 years following the switch. The 95% confidence interval is represented by the bars around the point estimate.

---

Figure 1 about here

---

The first panel represents total medical expenditure for Aetna Wellness members compared to non-members. Total medical expenditure increases for Aetna Wellness members over years of enrollment, with the increase starting immediately in the year of enrollment. Emergency expenditure does not show a clear trend in the first 4 years of enrollment, but may be inching up 6 to 7 years after enrollment. Pharmacy expenditure remains steady after enrollment in Aetna Wellness, and shows a decrease 6 to 7 years after enrollment. Expenditure on acute care does not show any trend over the length of enrollment. None of the trends, however, are statistically significant with large 95% confidence intervals particularly in the longer term enrollment period.

The event study suggests that overall expenditure may increase over time, with no decrease seen in emergency or acute expenditure. The difference-in-differences specification (3) is presented in Table 3 for the entire sample. Each row in the table is an estimate of equation (3) with a different dependent variable. Only the treatment ( $\beta_1$  coefficient on Aetna Wellness) and the after-treatment ( $\beta_2$  coefficient on Aetna Wellness \_ After) are reported in columns (1) and (2). Thus, the coefficients interpreted here will be those in column (2). Column (3) is the sample mean of the dependent variable.

---

Table 3 about here

---

Total expenditure increases by \$688 per person per year, but the increase is not statistically significantly due to the large standard errors. Aetna Wellness members, however, are likely to have lower expenditure before they enroll in the plan, at about \$1071 lower than non-members. This suggests that Aetna Wellness members are healthier or are lower spenders before enrollment. This also justifies the net decrease in expenditure seen in observational studies of wellness participants. The increase in overall expenditure is reflected in a comparable increase of \$639 in other medical expenditures, which includes outpatient care. Emergency and acute care expenditure does not change significantly to offset the rise in outpatient care.

The same pattern emerges when considering utilization. Looking at the total number of visits, the only significant change for Aetna Wellness members appears to be in outpatient care, in which preventive care is included. Outpatient visits increase by 1.57 visits per person per year, approximately a 20% increase on the average number of outpatient visits, of which preventive care accounts for

0.45 visit increase. Emergency and inpatient visits do not change significantly.

To explore whether Aetna Wellness members are less likely to seek medical care, the last section of Table 3 looks at any visits per year. Aetna Wellness members are 3 percentage points less likely to seek emergency care, but 13 percentage points more likely to have a preventive care visit. Finally, to estimate the availability of counseling and wellness features on their uptake, the last two specifications look at nutritional counseling, which Aetna Wellness members are 2 percentage points less likely to have, and mammography visits, which are not significantly different.

Four years may not be long enough for the health benefits of the continued preventive care to accrue in medical expenditure. Table 4 estimates the same specifications limiting the analysis on utilization 6 and 7 years after continued enrollment in Aetna Wellness. The smaller sample in this specification makes inference more difficult, however the magnitude and direction of the effects are consistent with the total sample results. Total medical expenditure increases by \$692, largely driven by a \$824 increase in other expenditure. Emergency and inpatient expenditure does not change significantly. Looking at the number of visits, as before, outpatient visits increase by 1.17 visits, of which preventive care account for 0.46, a significant increase. Aetna Wellness members remain more likely to seek preventive care, with no decrease in emergency care. Over the longer period, Aetna Wellness members are not more likely to take advantage of the nutrition counseling or mammography features of the plan.

Women make up a larger share of Aetna Wellness members, and may be driving the changes in utilization. Table 5 re-estimates the specification for women only. The smaller sample increases the standard errors making inference difficult, but the magnitude of the effects is comparable to those seen in Table 3. Most notable is the coefficient on mammogram, which is positive but not significant. Women may be taking greater advantage of the mammography features of Aetna Wellness, compared to non-members.

## 8 Discussion

The goal of Aetna Wellness was to increase use of preventive care to improve the health of its members and to facilitate diagnosis and maintenance of chronic conditions. As maintenance of health and chronic conditions improved, members would incur fewer emergency room and inpatient visits at the expense of increased preventive and outpatient visits. The results suggest that while the uptake of preventive care was successful, this was not offset by a parallel decline in emergency and inpatient acute care. Aetna Wellness members are more likely to use preventive care, and use it more often, though they are only slightly less likely to use the emergency room or inpatient care.

While 4 years might not be sufficient time for health gains of preventive care to materialize, the results of the longer-term users sample suggest that the pattern of utilization is persistent. After 6 and 7 years of preventive care and maintenance, Aetna Wellness members do not appear to use the emergency room or inpatient care less than non-members. Total medical expenditure continues to increase in the long term, along with outpatient and preventive care. Furthermore, while many of the features of a wellness plan are intended for both male and female members, there is no disproportionate tendency for either gender to use these features. Women are as likely to increase use of preventive care services as men.

According to the Bureau of Labor Statistics, the median number of years that wage and salary workers remain with their current employer was 4.6 years in 2014. The results presented here suggest that any increase in medical expenditure due to the preventive care will not be recuperated by the employer in future decreases in medical utilization.

Among the limitations of this study is the small sample size of the long term users. While the sample size limits our ability to draw inference from the estimates, the magnitude and direction

of the effects remain consistent across the larger and smaller sample, inspiring confidence in these results.

The single employer of this study is another limitation, as a university employee is not representative of the average American employee. As such, the results discussed here cannot be generalized. However, the limited availability of insurance choice and utilization data make these results important and significant. A larger study involving a more diverse group of employer across many industries will improve our understanding of the role of wellness plans in health insurance and utilization of medical services.

While a result presented here show a slight, though not significant, increase in total medical expenditure, the literature discussed in the previous sections shows total medical expenditure unchanged or declining. The difference in results can be attributed to multiple factors. The results presented here are limited to one employer in the education industry, as previously discussed.

The more significant source of difference between the current study and those cited above is the estimation method. The difference-in-differences estimation adds to the propensity score matching used by other studies to isolate the effect of the wellness features. As our results show, those who enroll in Aetna Wellness have lower medical expenditure than non-members, which would result in an overall decrease in medical expenditure in an observational setting.

Another source of the difference could be the design of the wellness program. The wellness program studied by Milani and Lavie (2009) not only offers the risk assessment and counseling, but it also offers workplace medical care by nurses. In the results here, the decline in hospital and emergency room is offset by increases in preventive care expenditure. If the workplace medical care replaces the preventive care, it would not be reflected in the health insurance utilization files and, therefore, underestimate the cost of preventive care in Milani and Lavie (2009).

Finally, the present analysis uses a longer period to evaluate the medical expenditure and utilization resulting from a wellness plan. Milani and Lavie (2009), by design compare medical expenditure 12 months prior and 12 months following the introduction of wellness plan. Henke et al. (2011) track 2 years of medical utilization data, which incorporates both the pre- and the post-period. Wellness programs may reduce utilization in the initial year, but utilization may return to pre-wellness levels in the subsequent years.

## **9 Conclusion**

The Aetna Wellness program was introduced in 2008 in order to improve the health maintenance, risk identification and prevention for employees and families by a large self-insured employer. While the features of the program were designed to help the employee, the increased emphasis on preventive care was intended to be reflected in the lower inpatient care and emergency room use. This study looks at the medical utilization of employees enrolled in the Aetna Wellness program, comparing their use to non-participants in the program.

The results suggest that there is no significant change in the overall medical expenditure, however, there is substantial compositional change in both expenditure and utilization by Aetna Wellness members. Those who switch into Aetna Wellness are more likely to engage in preventive and follow up outpatient care. The increase in preventive care does not lead to lower emergency or inpatient care in the medium or long term. Medical expenditure continues to be higher for Aetna Wellness members even 6 to 7 years after continued enrollment in the plan.

These results suggest that while wellness plans increase the use of preventive and outpatient care, they are not paralleled with savings in other medical utilization. Therefore, our results suggest that in the net wellness plans increase the total cost of an insurance plan to an employer.

## References

- Aldana, S. G., Merrill, R. M., Price, K., Hardy, A. & Hager, R. (2005), 'Financial impact of a comprehensive multisite workplace health promotion program', *Preventive medicine* **40**(2), 131–137.
- Cawley, J. & Price, J. A. (2013), 'A Case Study of a Workplace Wellness Program That Offers Financial Incentives for Weight Loss', *Journal of Health Economics* **32**(5), 794–803.
- Danagouliau, S. (2014), 'The Hassle of Wellness: Do Peers and Health Status Matter?', *Cornell University Dissertation* (Unpublished Material).
- Downey, M. (2014), 'The Doctor Is in Charge: How the ACA Puts the Employee's Physician in Charge of the Wellness Program', *Journal of Policy Analysis and Management* **33**(3), 820–826.
- Ehrlich, I. & Becker, G. S. (1972), 'Market Insurance, Self-Insurance, and Self-Protection', *Journal of Political Economy* **80**(4), 623–648.
- Grossman, M. (1972), 'On the Concept of Health Capital and the Demand for Health', *Journal of Political Economy* **80**(2), 223–255.
- Grossman, M. (2000), *The Human Capital Model*, Handbooks in Economics, vol. 17. Amsterdam; New York and Oxford: Elsevier Science, North-Holland.
- Grossman, M. & Rand, E. (1974), Consumer Incentives for Health Services in Chronic Illness, in S. Mushkin, ed., 'Consumer Incentives for Health Care', Milbank Memorial Fund, New York, pp. 114–151.
- Heckman, J. J., Ichimura, H. & Todd, P. (1998), 'Matching As An Econometric Evaluation Estimator', *The Review of Economic Studies* **65**(2), 261–294.

- Heckman, J. J., Ichimura, H. & Todd, P. E. (1997), 'Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme', *The Review of Economic Studies* **64**(4), 605–654.
- Henke, R. M., Goetzel, R. Z., McHugh, J. & Isaac, F. (2011), 'Recent experience in health promotion at Johnson & Johnson: lower health spending, strong return on investment', *Health affairs (Project Hope)* **30**(3), 490–499.
- Imbens, G. W. (2015), 'Matching Methods in Practice: Three Examples', *Journal of Human Resources* **50**(2), 373–419.
- Imbens, G. W. & Rubin, D. B. (2015), *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*, Cambridge and New York: Cambridge University Press.
- Kenkel, D. S. (2000), Prevention, in A. J. Culyer & J. P. Newhouse, eds, 'Handbook of health economics. Volume 1B', Vol. Handbook of Health Economics. Volume 1B, Handbooks in Economics, vol. 17; Amsterdam; New York and Oxford; Elsevier Science, North-Holland, pp. 1675–1720.
- Leuven, E. & Sianesi, B. (2015), 'PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing'.
- Marcus, J. (2014), 'Does Job Loss Make You Smoke and Gain Weight?', *Economica* **81**(324), 626–648.
- Merrill, R. M., Hyatt, B., Aldana, S. G. & Kinnersley, D. (2011), 'Lowering employee health care costs through the Healthy Lifestyle Incentive Program', *Journal of public health management and practice : JPHMP* **17**(3), 225–232.



- Milani, R. V. & Lavie, C. J. (2009), 'Impact of worksite wellness intervention on cardiac risk factors and one-year health care costs', *The American Journal of Cardiology* **104**(10), 1389–1392.
- Mukhopadhyay, S. & Wendel, J. (2013), 'Evaluating an Employee Wellness Program', *International Journal of Health Care Finance and Economics* **13**(3-4), 173–199.
- Naydeck, B. L., Pearson, J. A., Ozminkowski, R. J., Day, B. T. & Goetzel, R. Z. (2008), 'The impact of the highmark employee wellness programs on 4-year healthcare costs', *Journal of occupational and environmental medicine / American College of Occupational and Environmental Medicine* **50**(2), 146–156.
- Osilla, K. C., Van Busum, K., Schnyer, C., Larkin, J. W., Eibner, C. & Mattke, S. (2012), 'Systematic review of the impact of worksite wellness programs', *The American Journal of Managed Care* **18**(2), e68–81.
- Pauly, M. V. (1986), 'Taxation, Health Insurance, and Market Failure in the Medical Economy', *Journal of Economic Literature* **24**(2), 629–675.
- Rosenbaum, P. R. & Rubin, D. B. (1984), 'Reducing bias in observational studies using subclassification on the propensity score', *Journal of the American Statistical Association* **79**(387), 516–524.
- Shavell, S. (1979), 'On Moral Hazard and Insurance', *Quarterly Journal of Economics* **93**(4), 541–562.
- Smith, J. & Todd, P. (2005), 'Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators? Rejoinder', *Journal of Econometrics* **125**(1-2), 365–375.

- Stave, G. M., Muchmore, L. & Gardner, H. (2003), 'Quantifiable impact of the contract for health and wellness: health behaviors, health care costs, disability, and workers' compensation', *Journal of Occupational and Environmental Medicine* **45**(2), 109–117.
- Tolley, G., Kenkel, D. & Fabian, R. (1994), *Valuing health for policy: An economic approach*, Chicago and London: University of Chicago Press.
- Yen, L., Schultz, A. B., Schaefer, C., Bloomberg, S. & Edington, D. W. (2010), 'Long-term return on investment of an employee health enhancement program at a Midwest utility company from 1999 to 2007', *International Journal of Workplace Health Management* **3**(2), 79–96.
- Zweifel, P. & Breyer, F. (1997), *Health economics*, Translation. New York and Oxford: Oxford University Press.

**Table 1: Summary Statistics: Treated v. Control**

	<b>Control</b>	<b>Treated</b>
	Mean / Std. Err.	Mean / Std. Err.
Age	36.74 (.10)	37.42*** (.181)
Female	.49 (.002)	.56*** (.004)
Dependents	1.80 (.008)	2.02*** (.015)
Employee Wage	72,326 (307)	76,323*** (527)
Faculty	.27 (.002)	.24*** (.004)
Charlson Index	.19 (.003)	.16** (.006)
Medical Utilization Intensity	1.13 (.007)	1.30*** (.014)
Total Medical Expenses	5214 (89)	5023 (126)
Emergency Expenses	106 (2.59)	76*** (3.86)
Inpatient Acute Expenses	867 (51)	576** (63)
Pharmacy Expenses	1187 (21)	1063** (31)
Other Expenses	3052 (53)	3307* (83)
N	28,118	9,683
n	7,009	1,497

The star indicate the significance of the T-test statistic comparing the average for the control and treatment groups.

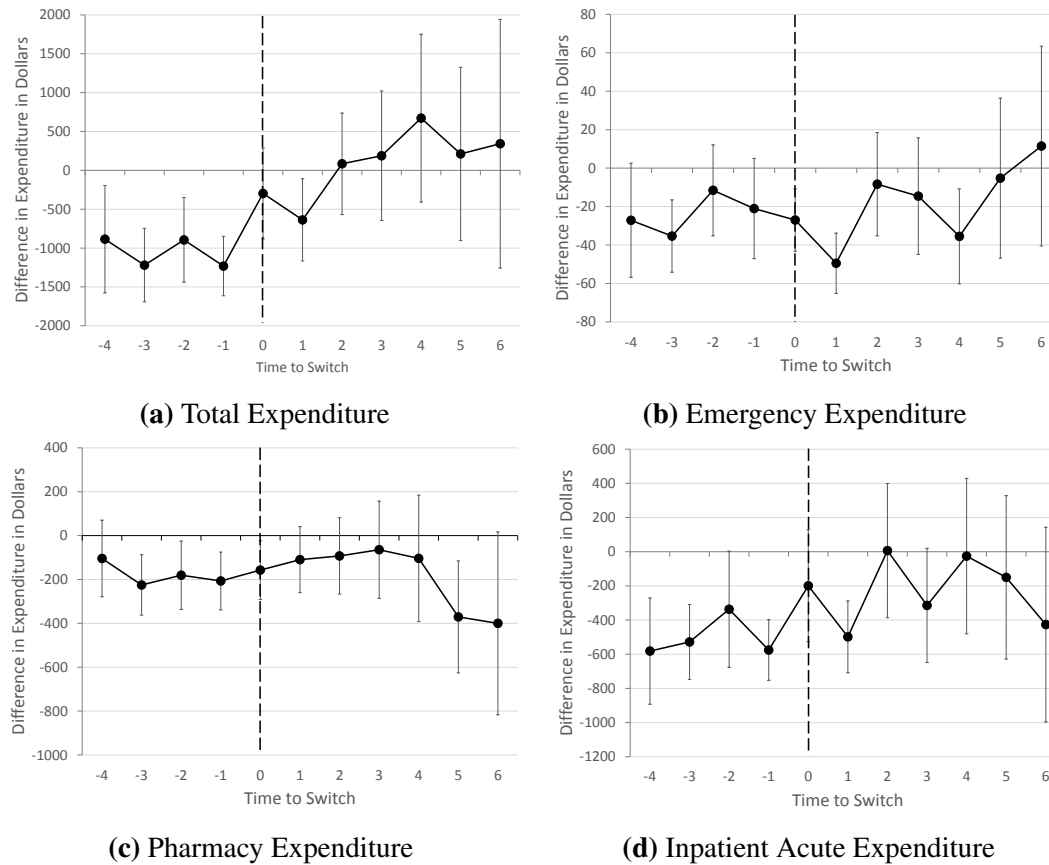
\*\*\* Significant at 0.1 percent level. \*\* Significant at 1 percent level. \* Significant at 5 percent level.

**Table 2:** Summary Statistics: Before and After Treatment

	<b>Before</b>	<b>After</b>
	Mean / Std. Err.	Mean / Std. Err.
Total Medical Expenses	4575 (241)	5287** (141)
Emergency Expenses	88 (7.1)	69* (4.4)
Inpatient Acute Expenses	570 (133)	579 (64)
Pharmacy Expenses	994 (60)	1103 (34)
Other Expenses	2922 (151)	3535*** (97)
N	3592	6091
n	1194	1194

The star indicate the significance of the T-test statistic comparing the average before and after.

\*\*\* Significant at 0.1 percent level. \*\* Significant at 1 percent level. \* Significant at 5 percent level.



**Figure 1:** Event Study by Expenditure Type<sup>1</sup>

<sup>1</sup>Each specification controls for individual health, insurance plan deductible for in-network care, gender, whether the person is a dependent, the number of dependents in the family, whether the employee is faculty or administrative staff, and race. While the sample is restricted to individuals who have been enrolled in Aetna Wellness at least 4 years, the estimates also track utilization in the years leading up to and following the switch, resulting in estimates for up to 4 years leading up to the switch, and up to 6 years following the switch. All treated observations are matched to control observations using propensity score. The 95% confidence interval is represented by the bars around the point estimate.

**Table 3:** Coefficient estimate of difference-in-differences model by medical expenditure and utilization type.

	(1) Aetna Wellness	(2) Aetna Wellness_After	(3) Dep. Mean
<i>Expenditure</i>			
Total Medical Expense	-1071.98*** (173.44)	688.15 (393.61)	5203.80
Emergency Expense	-29.39** (9.18)	-12.85 (12.87)	104.20
Pharmacy Expense	-184.61** (60.62)	-88.09 (122.77)	1181.35
Inpatient Acute	-530.29*** (100.33)	149.49 (173.52)	807.40
Other	-327.70** (102.39)	639.59* (292.39)	3110.84
<i>Number of visits</i>			
Emergency	-.06*** (.02)	-.04 (.02)	.24
Preventive	.09** (.03)	.45*** (.05)	.97
Inpatient	-.12*** (.02)	.04 (.03)	.17
Outpatient	.36 (.25)	1.57*** (.35)	8.12
<i>Any visits per year</i>			
Emergency	-.02* (.01)	-.03* (.01)	.17
Preventive	.05*** (.01)	.13*** (.02)	.57
Nutrition	.00 (.01)	-.02* (.01)	.09
Mammogram	.02** (.01)	.02 (.01)	.15

\*\*\* Significant at 0.1 percent level. \*\* Significant at 1 percent level. \* Significant at 5 percent level.

All coefficients represent the separate specification with the dependent variable specified in the first column.

The dependent variables are: Total - total allowed amount by insurance; ED - total allowed expenditure where emergency department is place of service; RX - total allowed pharmacy expenditure; IP Acute - total allowed inpatient acute medical expenditure; Other - all other medical expenditure; Emergency - number of visits where emergency department is the place of service; Preventive - number of preventive visits; Inpatient - number of inpatient visits; Outpatient - number of outpatient visits; Any Emergency - any emergency visits; Any Preventive - any preventive visits; Nutritional - any use of nutrition counseling; Mammography - any mammography visits.

The demographic variables are: age, gender, number of dependents, employee wage, Charlson index, medical utilization intensity, employee job family, race, and employee business unit. All specifications include year fixed effects.

Estimates are matched using kernel propensity score and standard errors clustered at individual level.

**Table 4:** Coefficient estimate of difference-in-differences model by medical expenditure and utilization type for long term users of Wellness.

	(1) Aetna Wellness	(2) Aetna Wellness_After	(3) Dep. Mean
<i>Expenditure</i>			
Total Medical Expense	-1307.68*** (370.87)	692.26 (624.10)	5289.01
Emergency Expense	-33.29* (13.21)	12.12 (20.58)	108.42
Pharmacy Expense	-350.08*** (83.18)	-163.43 (169.53)	1202.85
Inpatient Acute	-508.21 (267.98)	19.27 (325.88)	861.88
Other	-416.10* (205.31)	824.30 (444.42)	3115.86
<i>Number of visits</i>			
Emergency	-.05 (.03)	-.00 (.03)	.25
Preventive	.05 (.06)	.46*** (.07)	.93
Inpatient	-.10 (.06)	-.02 (.06)	.18
Outpatient	.71 (.53)	1.17 (.66)	8.00
<i>Any visits per year</i>			
Emergency	-.01 (.02)	-.01 (.02)	.18
Preventive	.02 (.02)	.16*** (.03)	.56
Nutrition	-.02 (.01)	-.01 (.02)	.09
Mammogram	.04** (.01)	-.00 (.02)	.15

\*\*\* Significant at 0.1 percent level. \*\* Significant at 1 percent level. \* Significant at 5 percent level.

All coefficients represent the separate specification with the dependent variable specified in the first column.

The dependent variables are: Total - total allowed amount by insurance; ED - total allowed expenditure where emergency department is place of service; RX - total allowed pharmacy expenditure; IP Acute - total allowed inpatient acute medical expenditure; Other - all other medical expenditure; Emergency - number of visits where emergency department is the place of service; Preventive - number of preventive visits; Inpatient - number of inpatient visits; Outpatient - number of outpatient visits; Any Emergency - any emergency visits; Any Preventive - any preventive visits; Nutritional - any use of nutrition counseling; Mammography - any mammography visits.

The demographic variables are: age, gender, number of dependents, employee wage, Charlson index, medical utilization intensity, employee job family, race, and employee business unit. All specifications include year fixed effects.

Estimates are matched using kernel propensity score and standard errors clustered at individual level.

**Table 5:** Coefficient estimate of difference-in-differences model by medical expenditure and utilization type for women.

	(1) Aetna Wellness	(2) Aetna Wellness_After	(3) Dep. Mean
<i>Expenditure</i>			
Total Medical Expense	-1164.27*** (225.10)	76.00 (468.29)	5307.59
Emergency Expense	-34.94** (10.83)	-20.78 (15.50)	102.31
Pharmacy Expense	-222.79** (85.26)	-173.92 (180.25)	1201.44
Inpatient Acute	-507.29*** (93.58)	-41.19 (199.29)	708.23
Other	-404.24** (146.87)	311.90 (324.61)	3295.61
<i>Number of visits</i>			
Emergency	-.06** (.02)	-.04 (.03)	.24
Preventive	.09* (.04)	.49*** (.06)	1.06
Inpatient	-.12*** (.02)	.01 (.03)	.15
Outpatient	.06 (.34)	1.54* (.49)	9.08
<i>Any visits per year</i>			
Emergency	-.02 (.01)	-.04* (.02)	.17
Preventive	.05** (.02)	.12*** (.02)	.62
Nutrition	-.01 (.01)	-.00 (.01)	.07
Mammogram	.04** (.01)	.03 (.02)	.28

\*\*\* Significant at 0.1 percent level. \*\* Significant at 1 percent level. \* Significant at 5 percent level.

All coefficients represent the separate specification with the dependent variable specified in the first column.

The dependent variables are: Total - total allowed amount by insurance; ED - total allowed expenditure where emergency department is place of service; RX - total allowed pharmacy expenditure; IP Acute - total allowed inpatient acute medical expenditure; Other - all other medical expenditure; Emergency - number of visits where emergency department is the place of service; Preventive - number of preventive visits; Inpatient - number of inpatient visits; Outpatient - number of outpatient visits; Any Emergency - any emergency visits; Any Preventive - any preventive visits; Nutritional - any use of nutrition counseling; Mammography - any mammography visits.

The demographic variables are: age, gender, number of dependents, employee wage, Charlson index, medical utilization intensity, employee job family, race, and employee business unit. All specifications include year fixed effects.

Estimates are matched using kernel propensity score and standard errors clustered at individual level.



# Appendix

## 1 First Order Condition

The first order conditions for the constrained maximization problem are

$$\frac{\partial I_t}{\partial M_t} \left[ Uh_{t+1} + (1 - \delta)Uh_{t+2} + \dots + (1 - \delta)^{n-1}Uh_n + \frac{\lambda g' W_{t+1}}{(1+r)^{t+1}} + \dots + \frac{\lambda(1 - \delta)^{n-1}g' W_n}{(1+r)^n} \right] + \frac{\lambda P_t}{(1+r)^t} = 0$$

$$\frac{\partial I_t}{\partial V_t} \left[ Uh_{t+1} + (1 - \delta)Uh_{t+2} + \dots + (1 - \delta)^{n-1}Uh_n + \frac{\lambda g' W_{t+1}}{(1+r)^{t+1}} + \dots + \frac{\lambda(1 - \delta)^{n-1}g' W_n}{(1+r)^n} \right] + \frac{\lambda V_t}{(1+r)^t} = 0$$

$$\frac{\partial I_t}{\partial TH_t} \left[ Uh_{t+1} + (1 - \delta)Uh_{t+2} + \dots + (1 - \delta)^{n-1}Uh_n + \frac{\lambda g' W_{t+1}}{(1+r)^{t+1}} + \dots + \frac{\lambda(1 - \delta)^{n-1}g' W_n}{(1+r)^n} \right] + \frac{\lambda W_t}{(1+r)^t} = 0$$

where  $Uh_t = \partial U / \partial H_t$ . Letting  $\pi_t$  represent the marginal cost of gross investment in health in period  $t$ , that is

$$\pi_t = \frac{P_t}{\partial I_t / \partial M_t} = \frac{V_t}{\partial I_t / \partial P V_t} = \frac{W_t}{\partial I_t / \partial T H_t}$$

the first order conditions reduce to

$$\frac{\pi_t}{(1+r)^t} = \sum_{i=1}^{n-t} (1 - \delta)^{i-1} \left[ \frac{Uh_{t+i}}{\lambda} + \frac{g' W_{t+i}}{(1+r)^{t+i}} \right] \quad (4)$$

## 2 Plan Choices

**Table A1:** Plan Comparison for a Family, All Plans 2012

	<b>Aetna Wellness</b>	<b>Aetna</b>	<b>Third</b>
Premium	3515	4555	3828
Deductible			
<i>In Network</i>	0	500	300
<i>Out of Network</i>	800	900	900
Out of Pocket Maximum			
<i>In Network</i>	4000	4100	4100
<i>Out of Network</i>	7000	7100	7100
Coinsurance			
<i>In Network</i>	90	90	90
<i>Out of Network</i>	80	80	70
Network	Aetna	Aetna	PHCS
Must choose PCP	Yes	No	No
Enhanced Wellness Program	Yes	No	No

## 3 Propensity Score Matching

The propensity score matching follows the steps outlined by Imbens and Rubin (2015) and Imbens (2015). The estimation is a two step process. In the first step, the propensity score is estimated using a logit specification:

$$Pr(T_{it}) = f(\text{Health}_{it-1}, X_{it}, t, \epsilon_{4,it}) \quad (5)$$

**Table A2: Summary Statistics Before and After Trimming and Balancing**

	Before		After	
	Control Mean / Std. Err.	Treated Mean / Std. Err.	Control Mean / Std. Err.	Treated Mean / Std. Err.
Age	36.96 (.10)	36.52* (.19)	37.60 (.11)	36.85** (.20)
Female	.50 (.002)	.56*** (.004)	.52 (.003)	.57*** (.005)
Dependents	1.80 (.008)	2.02*** (.015)	1.82 (.008)	2.05*** (.016)
Employee Wage	72,241 (304)	76,469*** (574)	74,100 (336)	75,707* (538)
Faculty	.27 (.002)	.23*** (.004)	.27 (.002)	.23*** (.004)
Charlson Index	.19 (.003)	.15*** (.006)	.19 (.003)	.15*** (.006)
Medical Utilization Intensity	1.14 (.007)	1.30*** (.015)	1.22 (.008)	1.30*** (.016)
N	33,882	8,816	27,152	8,276

The star indicate the significance of the T-test statistic comparing the average for the control and treatment groups.  
 \*\*\* Significant at 0.1 percent level. \*\* Significant at 1 percent level. \* Significant at 5 percent level.

where  $T_{it}$  is a treatment selection indicator,  $X_{it}$  is a vector of covariates,  $t$  are year fixed effects, and  $\epsilon_{4,it}$  is individual specific error. In the second step, the sample is trimmed to the overlap region, which excludes extreme values from the control group. This includes values with a propensity score well below and well above the treated sample. In the final step, the sample is divided into blocks, and the treated and control groups are compared based on the propensity score. If the difference between these is significant, the block is iteratively split until the groups are balanced.

This three step procedure allows the sample to be approximate the two requirements of unbiasedness for matching: the conditional independence assumption and the overlap assumption. Table A2 reports the covariate means for the treated and control before and after balancing. While the



**Figure 2:** Covariate by treatment status after balancing.

samples remain substantially different, the differences are reduced. In particular, the proportion of females in the control sample has increased, while the averages of wages for both treated and control have converged. The average intensity of medical utilization is higher for the control group, reducing the difference though it remains statistically significant. The mixed success of the propensity score matching requires difference-in-differences estimation structure to remove the selection bias in the treated group.