The Impact of Financial Incentives on Health and Healthcare: Evidence from a Large Wellness Program*

Liran Einav, Stephanie Lee, and Jonathan Levin†

May 2017

Abstract. Workplace wellness programs have become increasingly common in the United States, although there is not yet consensus regarding the ability of such programs to improve employees' health and reduce healthcare costs. In this paper, we study a program offered by a large US employer that provides substantial financial incentives directly tied to employees' health. The program has a high participation rate among eligible employees, around 80%, and we analyze data on the first four years of the program, linked to healthcare claims. We document robust improvements in employee health, and a correlation between certain health improvements and reductions in healthcare cost. Despite the latter association, we cannot find direct evidence causally linking program participation to reduced healthcare costs, although it seems plausible that such a relationship will arise over longer horizons.

*We thank Shirley Yarin for terrific research assistance, David Chen for helping us with the data, and Kent Bradley and Patricia Lin for answering our many questions. We acknowledge support from the Stanford Institute of Economics and Policy Research.

†Department of Economics, Stanford University (Einav, Lee, and Levin), Graduate School of Business, Stanford University (Levin) and NBER (Einav and Levin). Email: leinav@stanford.edu, slee88@stanford.edu, jdlevin@stanford.edu.

1 Introduction

Workplace wellness programs have become an important tool for insurers, employers, and policymakers to combat rising healthcare costs. These programs stem from the idea that encouraging working individuals to adopt a healthier lifestyle is a "win-win" strategy. Wellness programs have the potential to lead to better health for the individuals involved, translate to improved labor productivity, and reduce healthcare costs which have increased far faster than inflation over the last several decades.

Although some form of wellness program is offered by the majority of large employers in the United States (see Mattke et al. (2013) for a comprehensive review), there is not yet clear consensus as to the impact of such programs, or as to which program design features are most effective.[1] While the potential for cost saving appears large (Bolnick et al., 2013) and some studies suggest that wellness programs are associated with cost savings (Ozminkowski et al., 2002; Caloyeras et al., 2014; Baicker et al., 2010), others do not find a significant effect on cost (Aldana et al., 2005) or worry that such programs shift costs onto less healthy employees (Horwitz et al., 2013). On March 23, 2010, President Obama endorsed a provision created to encourage employers to implement more substantial standard-based wellness programs.

In this paper, we add to the existing evidence by exploiting rich administrative data from a large US employer which has been a pioneer in designing and implementing a workplace wellness program. The employer introduced its wellness program in late 2008. The program is innovative in its design and offers large financial incentives. A central feature of the program is that financial incentives are directly linked to employees' health. Through its workplace wellness program, the employer provides its employees strong financial incentives to pass specific health standards.

We describe the employer's workplace wellness program and the financial incentives, as well as our data, in Section 2. The program tracks five common health measures once a year – Body Mass Index (BMI), blood pressure, cholesterol, glucose, and nicotine. Our data contain administrative records for all eligible employees (and their program eligible dependents), as well as annual health measures for all program participants, and health insurance claim-level data for most individuals. We use data from the first four years of the program (2009-2012), covering more than 20,000 participants per year. Participation rates among eligible individuals are higher than in most other

[1] 71% of large firms with 200 or more employees with health benefits offer programs to help employees stop smoking, 68% offer lifestyle or behavioral coaching, and 61% offer programs to help employees lose weight. 81% of large firms report offering at least one of these three programs (Claxton et al., 2015).

programs. Perhaps the most unique feature of the program is its incentive structure. While most employers offer participation payments for employees, the employer we examine ties financial incentives directly to biometric measurements, offering up to $1,000 per year for individuals who do well on all metrics. The financial rewards (paid in the form of insurance premium discounts) can be eliminated entirely for individuals who fail all screenings.

In Section 3 we use a variety of empirical approaches to document a robust year-to-year improvement in each measured biometric for program participants. This complements findings reported in Fu et al. (2016) for the same employer.[2] There are two empirical challenges in attempting to attribute causal interpretation to the findings described above. First, participation in the program is voluntary, although internal advertising, benefits sessions, and the financial incentives lead to relatively high participation rates of 75-85% among eligible individuals. Second, health measures are only available for program participants during their participating years, so cannot be assessed against non-participants. While there is no perfect way to circumvent these challenges, we use external data to provide some benchmarks for comparison, which generally supports a causal interpretation.

Section 4 relates these health improvements to healthcare cost and utilization. Using a simple difference-in-differences specification, we do not find clear evidence that program participation is associated with lower healthcare cost, at least over the first four years of the program. In fact, we find some evidence for higher healthcare cost over the first two years. However, we do present evidence that health improvements in terms of BMI and blood pressure (but not in terms of cholesterol and glucose) are associated with lower healthcare utilization and cost, as well as suggestive evidence that the program triggers individuals to start treating (or adhere better) to their blood pressure and cholesterol reducing drugs.

In the final section we summarize our findings and conclude. Overall, while we do not find clear evidence that healthcare cost has declined as a result of the wellness program, our overall conclusion is quite positive. Participating employees appear healthier, and this will likely make them more productive and perhaps even cheaper in the longer run, regardless of the shorter run cost-benefit analysis.

[2] A related finding is reported by Cawley and Price (2013), who document a response by employees to financial incentives associated with weight loss.

2 Setting and Data

The Employer's Workplace Wellness Program The employer's workplace wellness program in its current form was first implemented for the calendar year 2009, and has continued throughout our sample period with small changes from year to year. All non-union employees are eligible for the program, as well as their spouses if they are covered (as dependents) by the employer's employer-provided health insurance. Participation in the program is voluntary, and has been around 75-85% among all eligible individuals.

As is typical in many employer-provided wellness programs, program participation requires individuals to take an annual confidential health screening and have its results reported to the program administration.[3] The screening can either be taken in a doctor's office, with the results transmitted to the company, or more commonly the program organizes and advertises pre-scheduled on-site events in which individuals could participate in the health screening session in their job location.

The health screening session takes about 15-30 minutes, and involves measuring five distinct health metrics, and a subsequent optional consultation with a health professional. The five health metrics are Body Mass Index (BMI), blood pressure, cholesterol, glucose, and smoking. Each is associated with a pass/fail outcome. Passing standards are based on standard health guidelines, with some leeway relative to the National Institutes of Health (NIH) recommendations.

BMI is the ratio of the individual's weight (in kilograms) to the square of her height (in meters), with a BMI of 30 or below considered a passing result.[4] Blood pressure is measured using systolic and diastolic mmHg readings, with passing result requiring that the reading is both below 140 (systolic) and below 90 (diastolic). Cholesterol is measured by total cholesterol, with values below 220 mg/dL considered a passing result.[5] To pass the glucose test (which started in 2010), glucose must be below 116 mg/dL, and to check for smoking individuals had to obtain a negative result on a nicotine test.[6]

[3] 50% of large firms with 200 or more employees with health benefits offer or require employees to complete an annual biometric screening (Claxton et al., 2015).

[4] Pregnant women automatically pass. Individuals with high BMI results could also pass the BMI metric by showing a 10% reduction in BMI in the subsequent year, even if the BMI is still above 30. Alternatively, individuals whose weight is high due to muscles buildup can elect to pass the BMI metric with waist circumference measure that is less than 40 (for males) or 35 (for females) inches.

[5] This is effective on 2011 and later. Cholesterol measurement did not take place in 2009, and in 2010 the requirements for a passing result was more stringent (HDL has to be greater than 40 mg/dL, LDL to be less than 130 mg/dL, and Triglyceride less than 200 mg/dL).

[6] The nicotine test result is binary, and the pass rate for those who take it is close to 100%, so we do not use this test throughout the analysis.

To the extent that financial incentives exist, the typical wellness programs provide such incentives by rewarding program participation. One of the unique features of the workplace wellness program at the employer we study is that the financial incentives to participants are large and are directly linked to successful test results. Each passing result on a biometric comes with a financial reward: a 3.00 to 5.50 dollar premium reduction in the *weekly* health insurance premium. These incentives can add up across five tests and a full calendar year to approximately 1,000 dollars per individual (or 2,000 dollars per household with two participating adults). Moreover, the program provides even stronger incentives to individuals who are less healthy. If an individual improves her health and meets the passing standard one year after having missed it in the prior year, she can receive a rebate for the measure retrospectively, so in such situations financial incentives are doubled. Appendix A provides complete details of the program rules.

Data Our data include annual information on all the eligible employees and spouses over five years (2008-2012), starting in the year before the start of the wellness program, and covering the first four years in which the program has been in place. In addition to the health measures that are available for all program participants, we also obtained administrative data on the employees' (and dependents) health insurance and pharmaceutical drug insurance claims for employees who are enrolled in the employer's PPO health insurance plan.

Table 1 presents summary statistics. Our initial, full sample – summarized in column (1) – consists of all individuals who were eligible for the workplace wellness program; that is, all non-union employees and their spouses who were covered by the employer's employer-provided health insurance. This sample consists of 115,805 individual-years that represent 41,590 unique individuals, of which 30,724 are employees and the rest are covered spouses. 46% of the observations are male, the average age is 45, and the majority have been working at the company for more than 10 years, with only 7.8% who were hired in 2009 or after. The employer has a national presence and operates across the country, but more than half of the sample contain individuals in California, Texas, and Oregon.

This initial sample has various elements of missing data for two primary reasons. First, health information is missing for program non-participants as their health measures are not being recorded. These account for 26% of the individual-years in the sample. Second, eligible individuals may opt to enroll in a health insurance plan that is different from the employer's PPO coverage. They are covered by Kaiser Permanente, which is a vertically-integrated healthcare provider. It receives capitated payments from the employer, and thus does not report back to the employer claim-level utilization information. These individuals account for 21% of the individual-years in the original sample.

We therefore end up working with different forms of samples, based on the specific analyses. The "health only" sample – summarized in column (2) of Table 1 – includes all participating individuals for whom we have health measures, the "utilization only" sample – summarized in column (3) – includes all PPO enrollees for whom we have utilization data, and the "complete data" sample (in column (4)) is the intersection of the two. As Table 1 shows, the samples are reasonably similar in terms of their observables (gender, age, salary, tenure at the company, and job location), although selection on unobservables (into the wellness program or into Kaiser coverage) is obviously a concern, which we will address below as well as we can.

3 Health changes over time

Health trends for program participants In this section we use the "health only" sample to explore how health changes for program participants. We start by looking at all individuals whose health is observed over consecutive years. That is, we drop individuals who are only in the sample for a single year or for non-consecutive years, and we treat each pair of consecutive years as a separate observation. Thus, for example, individuals who are observed for all four years are used to generate three different observations.

Table 2 shows the pass fail transition matrix for all consecutive year observations. As expected, health is highly correlated over time within an individual. Individuals who pass a given test in one year are much more likely to pass it in the subsequent year, with pass rates ranging from 91 to 94 percent conditional on passing the same test in the previous year, while pass rates are much lower – 18 percent for BMI and 45-65 for the other metrics – for those individuals who did not pass in the previous year.

Interestingly, Table 2 points to likely improvements in health metrics over time, as the transition matrix is not symmetric: the rate by which individuals transition from pass to fail is significantly lower than the rate by which they transition from fail to pass. To see this more granularly, Figure 1 presents the relationship between the health measure in a given year against the corresponding health measure of the same individual in the previous year. Across all measures, one can observe a clear pattern: individuals with high (worse) health metrics tend to have lower ones the year after, and individuals with low health metric tend to have higher ones. It is important to note that high measures are bad, but conditional on being low enough, lower measures are not necessarily better. In this sense, Figure 1 shows a clear pattern of improvements in health. To see this, note that we also plot the "passing threshold" in each panel of the figure, and in all panels the pattern crosses the 45-degree line below the passing threshold (and often even below the more stringent NIH recommendation), implying that individuals whose health metrics are too high tend to improve on average.

The pattern described above suggests that individuals who participate in the program tend to get healthier, at least on average, which is encouraging. Of course, this pattern doesn't mean necessarily that this improved health is caused by the program participation. The improvement in health may be viewed as even more encouraging once one realizes that aging alone "should" make many health metrics deteriorate from year to year, rather than improve. An important concern, however, is that the pattern observed in Figure 1 merely reflects mean reversion, either in true health or in the measurement of health. For example, a similar qualitative pattern could be generated if health is measured with error or if one's day-to-day health fluctuates (so a measurement in a particular day is a noisy signal of one's average daily health over the year).

To explore the potential importance of mean reversion and to more generally compare the pattern observed by the employer we study to some other benchmarks, we have searched for other data sets that follow individuals' health over time, and found two publicly available data sets that may serve as an imperfect benchmark. One is based on the Framingham Offspring Study and the other on the Coronary Artery Risk Development in Young Adults (CARDIA) study. The Framingham and CARDIA data sets are longitudinal studies designed to examine factors that influence the development of cardiovascular disease. Neither data set follows individuals annually, but over a longer time interval. The Framingham study has completed nine clinical examinations with intervals of 4-6 years between consecutive measures, while the CARDIA study has completed seven clinical examinations with intervals of 2-5 years between consecutive measurements. Appendix B provides more details about these data sets.

Figure 2 reports changes in BMI for Framingham, CARDIA, and workplace wellness program participants. For Framingham Offspring, changes in BMI are calculated as BMI from clinical examinations held in 1991-1995 minus BMI from clinical examinations held in 1987-1991. For CARDIA, changes in BMI are calculated as BMI from clinical examinations held in 2000-2001 minus BMI from clinical examinations held in 1995-1996. For workplace wellness program participants, changes in BMI are calculated as BMI from program year 2012 minus BMI from program year 2009. While the populations are not fully comparable, and the gap between measurement for the employer is slightly shorter, the pattern that emerges from Figure 2 is quite clear. It shows that the workplace wellness program participants are more likely to reduce their BMI than Framingham Offspring participants or CARDIA participants. CARDIA participants generally increase their BMI by 1 to 1.5 between two examinations regardless of their BMI from previous examination, presumably due to aging. The Framingham data shows some convergence (lower changes to BMI for individuals whose previous measures were higher), but still shows (as in the CARDIA data) that BMI on average increases regardless of the previous BMI measure. The employer data shows smaller changes overall, and in particular shows that when BMI from a previous examination (three years earlier) is greater than 30, The workplace wellness participants tend to reduce weight, whereas Framingham Offspring participants and CARDIA participants gain weight.

Differential health trends by participation duration An important limitation in our analysis is that we only observe health measures for individuals who participate in the program. Therefore, even if program participation was exogenous – and it is obviously unlikely to be the case – we do not have a way to compare health outcomes to a control group of non-participants, as for these individuals health measures are not observed. As a way to come closer to a comparable control group, in this section we compare pass rates of individuals who participated in the program for four years to pass rates of those who newly became eligible and participated in the program for the first time in 2012, the last year in our data. The idea behind the empirical strategy is that employees who were eligible for four years and employees who newly became eligible are comparable – both represent individuals who elected to participate in the program – except that the former group were exposed to the program for a significantly longer period (four years relative to only one). Furthermore, we can try to control for employee's decision to participate by looking at employees who have always participated to newly eligible employees who participated, so the comparison is across individuals who always participated in the program when they were eligible.

Table 3 cross tabulates the number of participation years against the number of eligibility years for individuals in our "complete data" sample. Our strategy is to compare health metrics in the "e1p1" cell – that is, individuals who were eligible only in 2012 and participated in the program in that year – to "e4p4" individuals who were eligible to participate in all four years (2009-2012) and participated in all years. There are 1,347 in the former group and 6,989 in the latter, and as shown in Table 3 it is notable that the full participation rates are similar: the 1,347 "e1p1" individuals constitute 56 percent of the individuals who were eligible, and the "e4p4" individuals constitute 53 percent of the individuals who were eligible in all years, perhaps suggesting that selection into participation in both groups is similar.

To take advantage of this variation, we compare pass rates for each health metric across the two groups, after controlling for year and age fixed effects. That is, we estimate the following regression

$$Pass_{it}^{m} = x_{it}'\gamma + \delta_t + \beta \cdot D_{e4p4} + \varepsilon_{it}, \tag{1}$$

where δ_t indicates year fixed effects, x_{it} is a vector of age fixed effects, and D_{e4p4} equals 1 if an individual was part of the group that was eligible for four years and participated for four years, and equals 0 if an individual was eligible for one year and participated for one year. We estimate this regression separately for each health metric, and separately for men and women. The results are reported in Table 4 and suggest that those who participated for four years are more likely to pass in 2012 than those who newly became eligible in 2012. The quantitative effects are quite large. Pass rates are approximately 5 percent higher for women across all metrics, and the differences are even larger among men for some of the metrics (e.g., 15 percent higher pass rates for cholesterol among men who participated all years).

Responses to changes in the magnitude of financial incentives In an alternative approach, we use the variation in incentive amounts to examine how individuals respond to financial incentives. The hypothesis is that people will respond more to greater incentives, and if they do, it seems likely that the program itself, which relies on financial incentives, improves health. In this analysis, we focus only on BMI and blood pressure because both these metrics were introduced in the first year of its workplace wellness program (2009) and testing standards remained unchanged throughout the observation period. Appendix Table A2 reports the year-to-year changes in financial incentives. The incentives depend on which PPO plan the individual was enrolled in, which we unfortunately do not observe. For both plans, the weekly incentive amounts have changed from year to year, both up and down, for administrative reasons that are unlikely to be associated with underlying health. Although the exact incentive amounts differ across the two plans, incentive amounts have changed similarly for both plans. To take advantage of this variation, we use the incentives associated with the higher coverage PPO plan ("Choice Fund I"), which covered more individuals. For individuals enrolled in this plan, passing BMI implied a benefit of 6 dollars per week in 2009 and 2010, but only 4 dollars in 2011, and 5.50 dollars in 2011. Similarly, passing blood pressure implied a benefit of only 1.50 dollars in 2009, 4 dollars in 2010 and 2011, and 3 dollars in 2012.

Using the variation in incentive amounts, we estimate

$$Pass_{imt} = \alpha_i + \beta \cdot WI_{mt} + \theta_m + \delta_t + \varepsilon_{imt}, \tag{2}$$

where the dependent variable is equal to one if an individual i passed health measure m in program year t. θ_m and δ_t are, respectively, health measure and year fixed effects, and α_i represents individual fixed effects. WI_{imt} is the weekly incentive amount associated with passing health measure m in year t, and β is the main coefficient of interest. We estimate this regression using all individual-year observations in the "complete data" sample. The average passing rate in the sample is 0.80 (across BMI and BP metrics), and we estimate an effect of 0.012 (standard error of 0.001). That is, every dollar in increased financial benefits (per week) is associated with a statistically significant 1.2 percentage points (approximately 1.5 percent) higher pass rates.

4 Associated changes in healthcare expenditure

An important motivation for wellness programs across the country is the premise that beyond the obvious and indirect benefits associated with healthier workforce, there are also potential direct benefits in the form of lower medical healthcare expenditure, which would more directly translate to cost savings for employers. In this section, we use the employer's healthcare claims to provide evidence that relates health measures to healthcare expenditure.

The basic fact that in the cross section healthier individuals spend, on average, less on healthcare than sicker individuals is widely documented,[7] and in Appendix Figure A1 we present a similar pattern in the context of our data. It may be less obvious that relatively short-run improvements in health translate to lower medical spending. To explore this in the context of our data, we estimate the following equation

$$\log(1 + MedExp_{it}) = \alpha_i + \delta_t + \beta \cdot Health_{it} + \varepsilon_{it}, \tag{3}$$

where $Health$ represents an individual health measure in a given year, and α_i and δ_t represent individual and year fixed effects. The dependent variable is the (logarithm of) total medical ex-

[7] See, e.g., Pronk et al. (1999), Sturm (2002), Wee et al. (2005), Finkelstein et al. (2009, 2010), and Hammond and Levine (2010).

penditure of individual i in year t, which we obtain by aggregating all of the individual's (who is covered by one of employer's PPO health plans) claims. We estimate this regression separately for each health measure, and the parameter of interest β is identified off variation in health measures for a given individual over time. Table 5 presents the results. Interestingly, the results suggest that reductions in BMI and blood pressure are associated with non-trivial reductions in healthcare spending, while changes in glucose and cholesterol measurements do not appear to systematically correlate with healthcare expenditure. For example, the estimates suggest that a one point reduction in BMI – which is approximately a three kilogram reduction in weight for most individuals – is associated with more than a 1.5% reduction in expected medical costs.

Moreover, because healthcare expenditure data is available for all employees (and their dependents) who enrolled in employer's PPO health plan, we can also report a "reduced form" estimate of the overall effect of the employer's workplace wellness program on healthcare expenditure. Specifically, we estimate the following regression

$$\log(1 + MedExp_{it}) = \alpha_i + \delta_t + \sum_{t=2009}^{2012} \beta_t \cdot Participation_{it} + \varepsilon_{it}, \tag{4}$$

where $Participation_{it}$ is a dummy variable that is equal to one if individual i participated in the program in year t and zero otherwise. We again use individual and time fixed effects. Identification is primarily driven by incorporating into the data observations on 2008 healthcare spending, at which point individuals have yet to participate in the program. We also obtain identification from the less frequent occasion in which eligible employees enroll in the program later or drop off. This variation is obviously imperfect, so any attempt to attribute causal interpretation to the estimated coefficients should be cautioned. With this caveat in mind, the results, presented in Figure 3, suggest that initially (2009 and 2010) program participation is in fact associated with increased healthcare expenditure, but later the healthcare expenditure declines to levels that are at or below the pre-program spending levels. One reasonable (in our view) interpretation of these results is that initially the workplace wellness program causes individuals to pay more attention to their health than before and incur certain healthcare expenditure, such as additional tests or preventive medication, that increases healthcare cost. However, in the longer run such increased expenditure improves one's health and individuals thus reduce their (curative) spending.

There are multiple ways by which an individual could go about improving health measures, ranging from lifestyle and dietary changes to taking prescribed medications. Of the four health measures that are being used in the program, two – blood pressure and cholesterol – have a very easy ramification: preventive medication (see, e.g., Krousel-Wood et al., 2004; Cholesterol Treatment Trialist, 2008). Blood pressure drugs are quite effective in reducing one's blood pressure and anti-cholesterol drugs are effective in reducing one's cholesterol level. Figure 4 shows the overall improvement in these two health measures during the course of the sample, with noticeable decline in the higher end of the spectrum. Consistent with preventive drugs playing a role in these improvements, Figure 5 plots the timing of blood pressure and anti-cholesterol drug purchases, separately for individuals who passed and failed the screening. The figure clearly shows that the timing of drug purchase is not at all associated with the screening date for those individuals who passed the screening, presumably because such individuals are taking care of their blood pressure and/or cholesterol condition regularly. Yet, for individuals who failed the screening, we see a clear pattern where they are much more likely to take preventive drugs after the screening relative to before, validating that the screening either alerts individuals to their previously undiagnosed condition or nudges them toward adherence.

5 Conclusions

In this paper we evaluated the impact of a novel wellness program of a large US employer. We find robust evidence that health biometrics improved for program participants, and that these improvements – at least for BMI and blood pressure – are associated with reduced healthcare cost and utilization. We do not find clear causal evidence for an overall cost saving.

As wellness programs have spread, there is increasing interest and debate about their efficacy. This paper adds an additional data point in attempting to assess whether these programs do in fact improve employee health and reduce healthcare costs. We do not find clear evidence for overall cost reduction; however, it is possible this objective should get less consideration in the short run. As we documented in Section 4, healthcare costs initially may rise with additional screening, and in fact we observe an increase in the use of preventive medication with the program. At the same time, preventive care may imply a healthier workforce, improved productivity, and eventually lower costs. Assessing the longer run impact of wellness programs is an important avenue for future work (although would require data from a different employer as the specific program we studied in this paper was recently discontinued).

References

Aldana, Steven G., Ray M. Merrill, Kristine Price, Aaron Hardy, and Ron Hager. 2005. "Financial impact of a comprehensive multisite workplace health promotion program." *Preventive Medicine* 40(2), 131-137.

Baicker, K., D. Cutler, and Z. Song. 2010. "Workplace Wellness Programs Can Generate Savings." *Health Affairs* 29(2), 304-311.

Bolnick, Howard, Francois Millard, and Jonathan P. Dugas. 2013. "Medical care savings from workplace wellness programs: what is a realistic savings potential?" *Journal of Occupational and Environmental Medicine* 55(1), 4-9.

Caloyeras, John P., Hangsheng Liu, Ellen Exum, Megan Broderick, and Soeren Mattke. 2014. "Managing Manifest Diseases, But Not Health Risks, Saved PepsiCo Money Over Seven Years." *Health Affairs* 33(1), 124-131.

Cawley, John, and Joshua A. Price. 2013. "A case study of a workplace wellness program that offers financial incentives for weight loss." *Journal of Health Economics* 32(5), 794-803.

Cholesterol Treatment Trialists. 2008. "Efficacy of cholesterol-lowering therapy in 18 686 people with diabetes in 14 randomised trials of statins: a meta-analysis." *Lancent* 371(9607), 117-125.

Claxton, Gary, Matthew Rae, Nirmita Panchal, Heidi Whitmore, Anthony Damico, Kevin Kenward, and Michelle Long. 2015. "Health benefits in 2015: stable trends in the employer market." *Health Affairs* 10-1377.

Finkelstein, Eric A., Justin G. Trogdon, Joel W. Cohen, and William Dietz. 2009. "Annual medical spending attributable to obesity: payer-and service-specific estimates." *Health Affairs* 28(5), w822-w831.

Finkelstein, Eric A., Marco daCosta DiBonaventura, Somali M. Burgess, and Brent C. Hale. 2010. "The costs of obesity in the workplace." *Journal of Occupational and Environmental Medicine* 52(10), 971-976.

Fu, P.L., K. Bradley, S. Viswanathan, J.M. Chan JM, and M. Stampfer. 2016. "Trends in Biometric Health Indices Within an Employer-Sponsored Wellness Program With Outcome-Based Incentives." *American Journal of Health Promotion* 30(6), 453-457.

Hammond, Ross A., and Ruth Levine. 2010. "The economic impact of obesity in the United States." *Diabetes, metabolic syndrome and obesity: targets and therapy* 3, 285.

Horwitz, Jill R., Brenna D. Kelly, and John E. DiNardo. 2013. "Wellness incentives in the workplace: cost savings through cost shifting to unhealthy workers." *Health Affairs* 32(3), 468-476.

Krousel-Wood, Marie, Sheila Thomas, Paul Muntner, and Donald Morisky. 2004. "Medication adherence: a key factor in achieving blood pressure control and good clinical outcomes in hypertensive patients." *Current Opinion in Cardiology* 19(4), 357-362.

Mattke, Soeren, Hangsheng Liu, John P. Caloyeras, Christina Y. Huang, Kristin R. Van Busum, Dmitry Khodyakov, and Victoria Shier. 2013. *Workplace Wellness Programs Study: Final Report.* RAND Corporation. Available at http://www.rand.org/pubs/research_reports/RR254.html.

Ozminkowski, Ronald J., Davina Ling, Ron Z. Goetzel, Jennifer A. Bruno, Kathleen R. Rutter, Fikry Isaac, and Sara Wang. 2002. "Long-term impact of Johnson & Johnson's Health & Wellness Program on health care utilization and expenditures." *Journal of Occupational and Environmental Medicine* 44(1), 21-29.

Pronk, Nicolaas P., Michael J. Goodman, Patrick J. O'connor, and Brian C. Martinson. 1999. "Relationship between modifiable health risks and short-term health care charges." *Journal of the American Medical Association* 282(23), 2235-2239.

Sturm, Roland. 2002. "The effects of obesity, smoking, and drinking on medical problems and costs." *Health Affairs* 21(2), 245-253.

Wee, Christina C., Russell S. Phillips, Anna TR Legedza, Roger B. Davis, Jane R. Soukup, Graham A. Colditz, and Mary Beth Hamel. 2005. "Health care expenditures associated with overweight and obesity among US adults: importance of age and race." *American Journal of Public Health* 95(1), 159.

Appendix A: Complete details of the Employer's Workplace Wellness Program

Eligible participants can participate in any or all of the health measures they choose. Receiving incentive discounts for one passing measure is not contingent upon participation or meeting standards for other measures. The program's details regarding pass standards and incentive structure have slightly changed during each program year.

Appendix Table A1 summarizes passing standards for each measure and year (2009-2012). BMI is the ratio of the individual's weight (in kilograms) to the square of her height (in meters), with a BMI of 30 or below considered a passing result. Alternatively, individuals can pass the BMI metric with waist circumference measure that is less than 40 (for males) or 35 (for females) inches. Blood pressure is measured using systolic and diastolic mmHg readings, with passing result requiring that the reading is both below 140 (systolic) and 90 (diastolic). In 2009, cholesterol testing was only required of those in a "high risk" demographic group. Those at "low risk" received the cholesterol incentive by filling out a documentation form. Cholesterol and Glucose measurements started in 2010, and passing standards for Cholesterol and Glucose have been relaxed by the employer in 2011. In 2010, individuals had to pass both Cholesterol and Glucose standards to receive financial incentives.

For each health measure participants passes, individuals receive financial incentives as shown in Appendix Table A2. Because the law regulates incentive amounts based on medical plan's costs, there are two different incentive schemes which depend on the individual's choice of PPO health plan. If an employee does not pass a test, they can receive incentives retroactively by re-testing and meeting the standard the following year. For those who do not meet the BMI standard, they can also reduce their body weight by 10 percent or more in the subsequent year and pass the screening tests (even if the new weight does not meet the standard).

In 2010, an auto-pass system was introduced where if individuals passed all five biometrics by satisfying the NIH recommended levels, which are more stringent than the employer's (see Figure 1), individuals will not have to test again in the subsequent year and will automatically receive financial incentives for the subsequent year. The auto-pass requirements were as follows: BMI between 18.5 and 24.9, blood pressure less than 120/80, total cholesterol less than 200, glucose less than 100, and negative nicotine test.

Health screenings take place during the fall open enrollment period, and the financial incentive is applied to the following year's health insurance premium. Employees are informed about screening opportunities via a home mailer and postings on the company intranet. Employees also have access to an online benefits portal where screening details, scheduling dates, and deadlines are accessible. Health screenings can be done through on-site workplace screening, submission of results through independent labs or personal physicians, or a testing kit mailed to the employee's home. A waiver can be submitted by individuals with special medical conditions in which it would be medically inadvisable or unreasonably difficult to satisfy the standard, such as pregnancy and Type 1 diabetes.

Appendix B: Additional details about the CARDIA and Framingham data sets

CARDIA. The Coronary Artery Risk Development in Young Adults Study (CARDIA) is a study examining the development and determinants of clinical and subclinical cardiovascular disease and its risk factors. It began in 1985-1986 with a group of 5,115 black and white men and women aged 18-30 years. The participants were selected so that there would be approximately the same number of people in subgroups (of race, gender, education, and age) in each of 4 centers: Birmingham, AL; Chicago, IL; Minneapolis, MN; and Oakland, CA. These same participants were asked to participate in follow-up examinations. The study has completed seven clinical examinations with intervals of 2-5 years between consecutive measurements. Exam 5 held in 1995-1996 and included 3,883 individuals. Exam 6 held in 2000-2001 and included 3,627 individuals. 3,322 individuals aged 32-49 (mean age 40) participated in both Exam 5 and Exam 6 to get their BMI measured in both exams. The changes in BMI from Exam 5 to 6 provide benchmarks to compare BMI patterns observed at the employer. Additional details are available at https://www.nhlbi.nih.gov/research/resources/obesity/population/cardia.htm.

Framingham. The objectives of the Framingham Offspring Study are to study the incidence and prevalence of cardiovascular disease (CVD) and its risk factors. The original Framingham study began in 1948 with 5,209 adult subjects from Framingham, MA. With the aging of the Framingham cohort and with interest in familiar aggregation and heritability, a new cohort consisting of the offspring of the original cohort was sampled. Spouses of offspring were also included. This new sample, begun in 1971, constituted a second generation of participants, permitted new assessment of risk factors and outcomes, and provided a resource for the genetic analyses which were yet to come. The Framingham Offspring Study include 5,124 men and women, ages 5-70 years at entry consisting of offspring of the original Framingham cohort. By 2014, the study has completed nine clinical examinations with intervals of 4-6 years between consecutive measures. Exam 4 held in 1987-1991 and included 3,903 individuals. Exam 5 held in 1991-1995 and included 3,683 individuals. 3,506 individuals aged 28-76 (mean age 51) participated in both Exam 4 and Exam 5 to get their BMI measured in both exams. The changes in BMI from Exam 4 to 5 provide benchmarks to compare BMI patterns observed at the employer. Additional details are available at https://biolincc.nhlbi.nih.gov/studies/framoffspring.

Appendix C: Additional details on the anti-cholesterol and blood pressure prescription drug claims analysis

We use individuals' prescription drug claims to describe the timing of blood pressure and anti-cholesterol medication claims. To identify blood pressure and anti-cholesterol medication, we used the list of common medications used to treat high blood pressure (http://www.webmd.com/drugs/condition-1432-High%20Blood%20Pressure%20%20Hypertension%20.aspx) and common medications used to treat high cholesterol (http://www.webmd.com/drugs/condition-701-High%20cholesterol%20%20 Hypercholesterolemia%20.aspx), respectively. We then use label names included in the prescription drug claims to identify blood pressure and anti-cholesterol medication.

The prescription drug claims data cover 1,948,794 drug claims made by 45,903 unique individuals (employees and their spouses) in 2008-2012. The drug claims include a total of 302,220 blood pressure drug claims made by 14,041 unique individuals and 139,092 anti-cholesterol drug claims made by 9,313 unique individuals.

We observe the date of each biometric examination and the date of drug purchase. To study the timing of blood pressure medication claims, we examine when individuals get their blood pressure purchased relative to their blood pressure examination date. We examine blood pressure drug purchases that took place between 240 days before and 240 days after the 2009 blood pressure examination. Observations are 36,897 blood pressure drug purchases made by 3,322 unique individuals who participated in blood pressure examination in 2009. Among these 3,322 individuals, 2,450 unique individuals passed blood pressure examination in 2009 and made 27,051 blood pressure drug purchases (240 days before and after the examination). The remaining 872 individuals failed blood pressure screening and made 9,846 blood pressure drug purchases.

To examine the timing of cholesterol drug medication claims, we examine cholesterol drug purchases that took place between 240 days before and 240 days after the 2011 cholesterol examination. Cholesterol measurements started in 2010, and cholesterol became an independent biometric starting 2012. (In 2010, individuals had to pass both Cholesterol and Glucose standards to receive financial incentives.) We observe 21,031 anti-cholesterol drug purchases (made between 240 days before and after the cholesterol examination) made by 2,611 unique individuals who participated in cholesterol screening in 2011. Among these 2,611 individuals, 2,168 individuals passed the screening and made 18,704 anti-cholesterol drug purchases (240 days before and after the examination), while the remaining 443 individuals failed the cholesterol screening and made 2,327 cholesterol corresponding drug purchases.

Figure 1: Year-to-year health measures test results

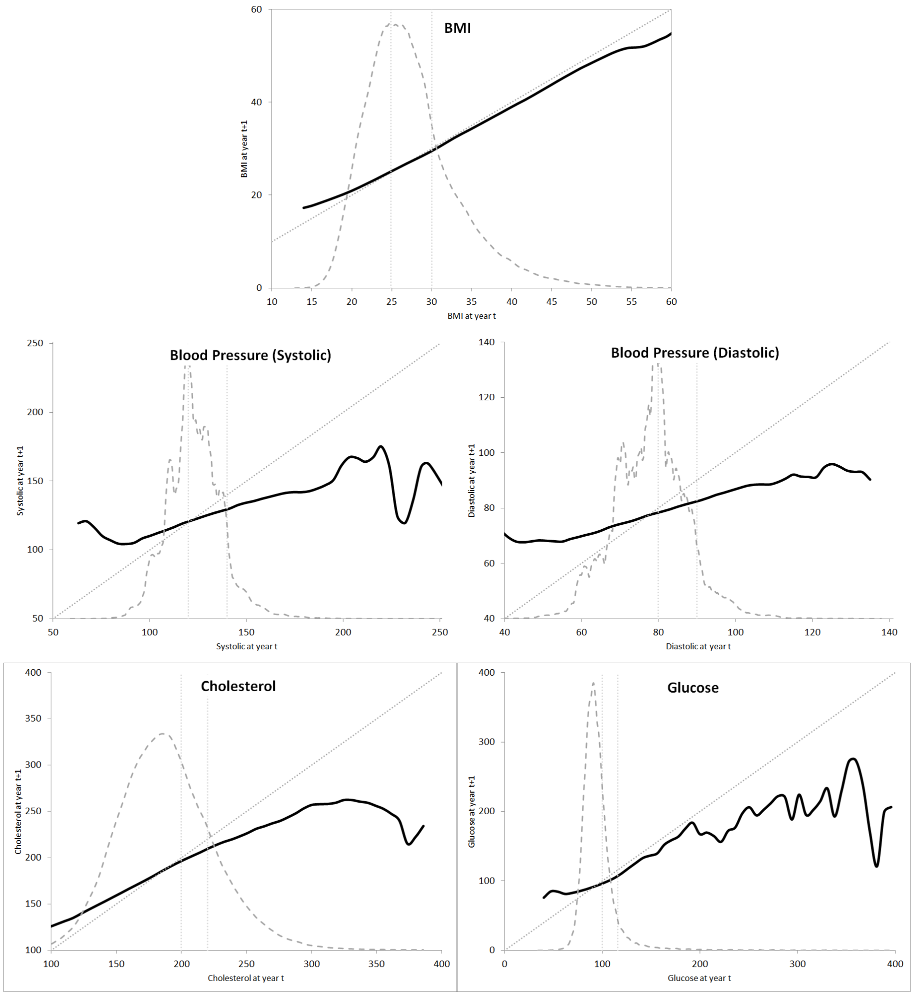

Figure plots (thick black lines) predictions from kernel regression of a health indicator in one year on the same health indicator the year before, using all available observations on the same person in consecutive years. Each panel presents one type of test. All panels show the 45-degree line, two vertical lines at the employer passing threshold and the (lower) NIH recommended level, and the underlying distribution of year t-1 values (dashed gray).

Figure 2: Comparison of workplace wellness participants to other populations

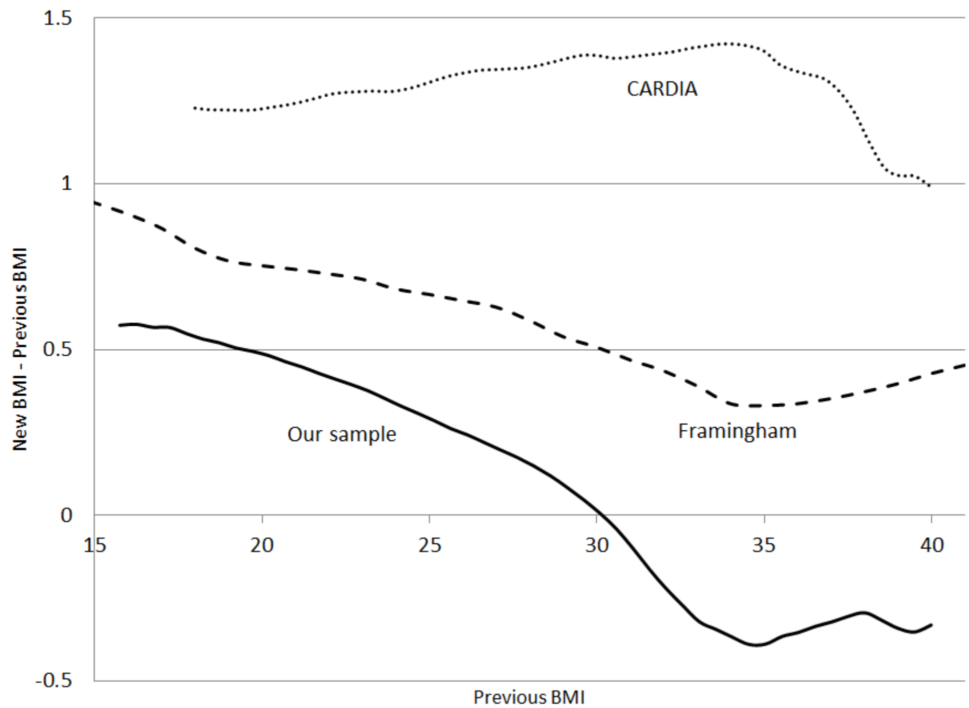

Figure shows the estimated change in BMI from one examination to the next examination, estimated using a local polynomial regression plotted against BMI from the previous examination. For CARDIA, previous BMI indicates BMI from examinations held in 1995-1996, and new BMI indicates BMI from examinations held in 2000-2001. For Framingham, previous BMI indicates BMI from examinations held in 1987-1991, and new BMI indicates BMI from examinations held in 1991-1995. For workplace wellness program participants, previous BMI indicates BMI from examinations held in program year 2009, and new BMI indicates BMI from examinations held in program year 2012. See Appendix B for more details on the CARDIA and Framingham data sets.

Figure 3: Relationship between program participation and healthcare expenditure

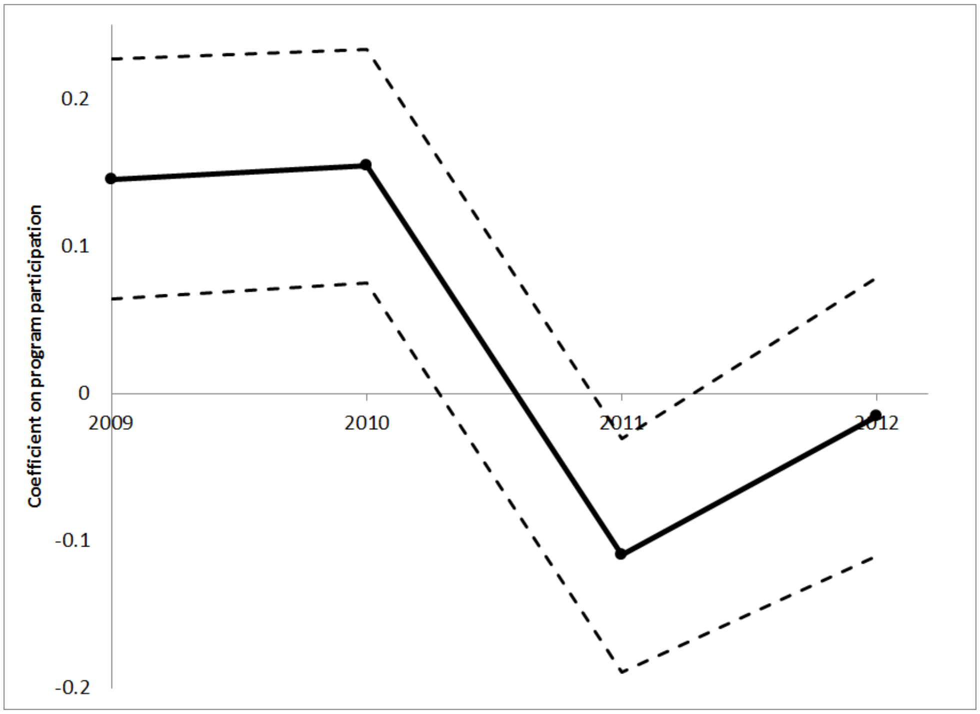

Figure plots estimated coefficients and 95 percent confidence interval of β_t from estimating equation (4). Observations are individual-years in the "utilization only" sample, but also include observations from year 2008, the last year before the program started. There is a total of 109,288 individual-years observations (18,135 from 2008 and 91,153 from 2009-2012), covering 32,939 unique individuals.

Figure 4: Improvements in blood pressure and cholesterol measures over time

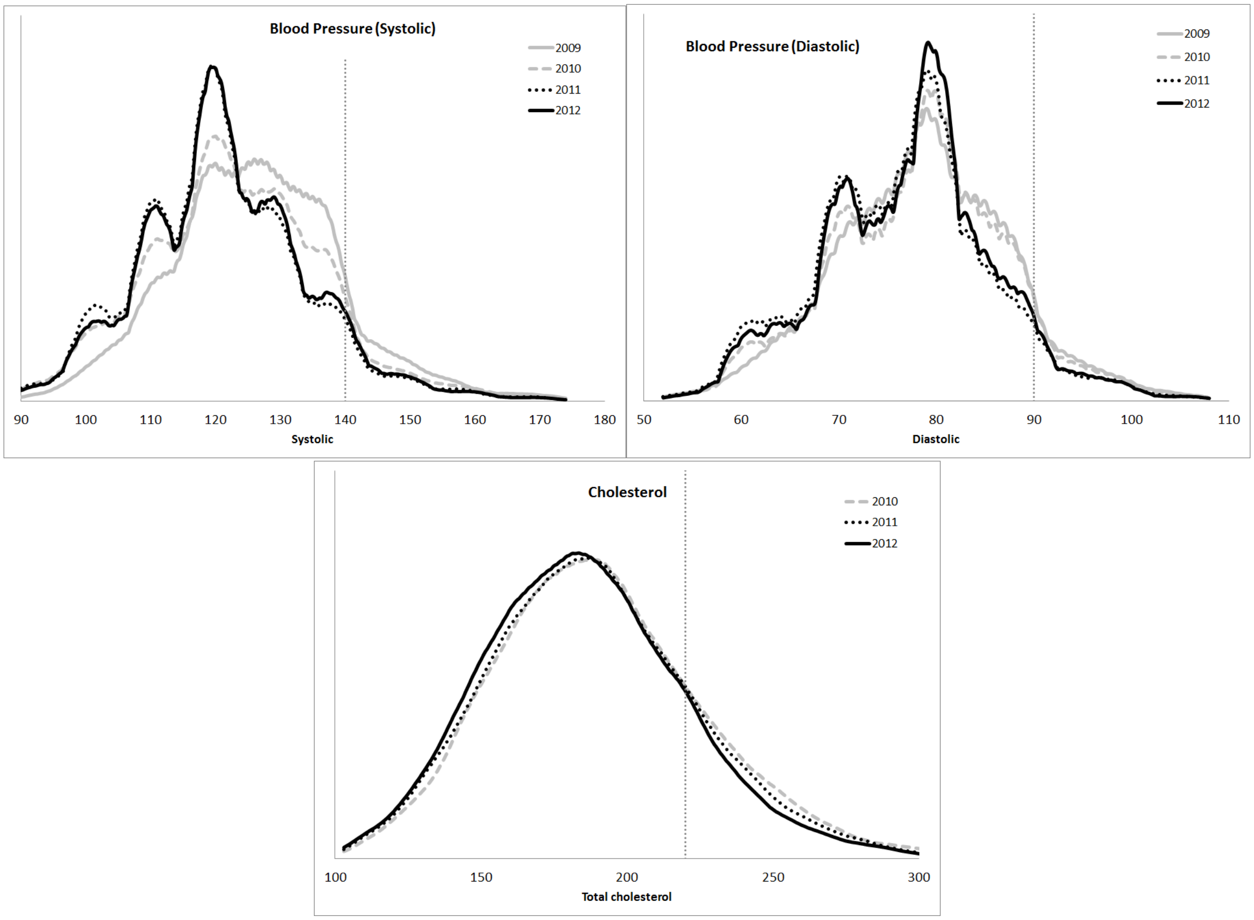

Figure plots kernel densities of the cross-sectional distribution of systolic (top left), diastolic (top right), and total cholesterol (bottom), for each year (recall that cholesterol screening began in 2010). Observations are all individual-years in the "health only" sample that were measured. Measurements are missing for some individuals who auto-passed by passing all five metrics by satisfying NIH standards in the previous year or received an exemption. We dropped observations below the 0.5th percentile and above the 99.5th percentiles. The vertical lines shows the passing threshold, above which an individual fails. The sample includes 77,958 individual-years observations (32,142 unique individuals) for systolic, 77,921 (32,116) for diastolic, and 54,892 (26,722) for cholesterol.

Figure 5: The timing of preventive drug purchasing and screening results

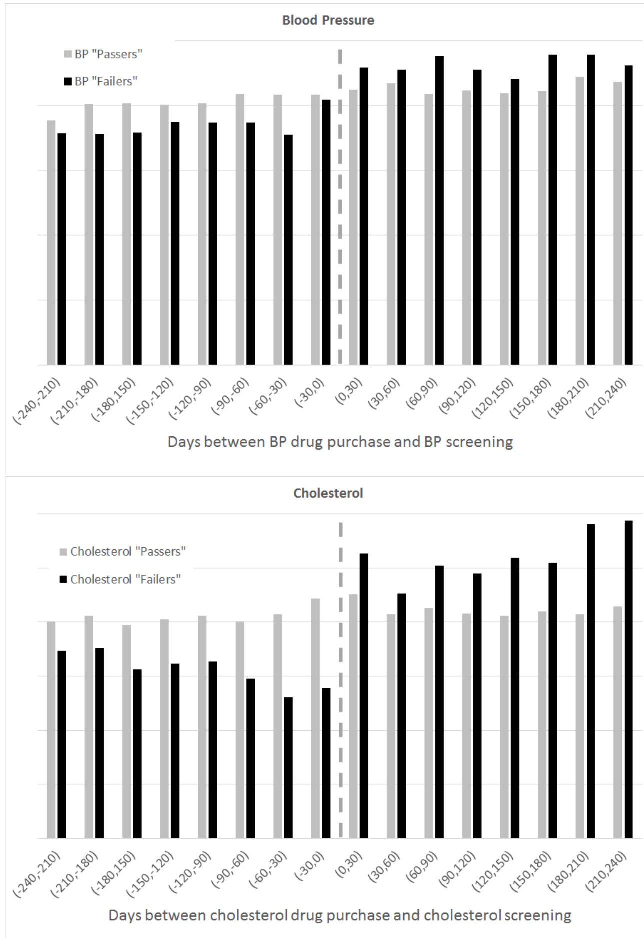

Figure presents the distribution of the timing of preventive drug purchasing around the screening date. The top panel shows results for blood pressure screening (and blood pressure medications) in 2009, and the bottom panel shows results for cholesterol screening (and anti-cholesterol drugs) in 2011. The key variable is the difference between the drug purchase and the screening date, with negative (positive) numbers reflect purchases that occurred before (after) the screening. Appendix C provides more details on the data construction. In both panels, the black bars present (normalized) frequency for those individuals who failed the screening, while the gray bars repeat the analysis for those who passed.

Table 1: Summary statistics

	Full sample (1)	"Health only" (2)	"Utilization only" (3)	"Complete data" (4)
Observations				
Individual-years	115,805	85,673	91,153	61,021
Unique individuals	41,590	34,212	32,939	24,640
Unique employees	30,724	26,059	24,030	18,667
Demographic Information*				
Male	0.46	0.45	0.46	0.45
Age^ (avergae)	45.15	44.74	45.86	45.56
Age^ (std. deviation)	11.78	11.75	11.53	11.30
Salary (average, in $000)	58.43	61.95	58.87	61.95
Salary (std. deviation)	37.32	38.20	37.14	38.20
Year of hire (average)	1997.6	1996.8	1997.3	1996.8
Share hired in 2009 or later	0.078	0.056	0.064	0.056
State of employment				
CA	0.23	0.23	0.19	0.19
TX	0.17	0.17	0.20	0.20
OR	0.16	0.16	0.16	0.16
Other	0.44	0.44	0.45	0.45
Health expenditure				
Annual Health Spending			5,117	4,865
Std. Dev.			19,668	17,911
Annual Drug (RX) Spending			789	947
Std. Dev.			3,159	3,079
Passing rates				
BMI		0.76		0.75
BP		0.85		0.85
Cholesterol (2010-2012)		0.69		0.69
Glucose (2010-2012)		0.57		0.56

Observations are from 2009-2012. "Health only" sample includes all individuals who participated in the program that year. "Utilization only" sample includes all individuals enrolled in one of the company's PPO options that year. "Complete data" sample includes individual-years who satisfy both criteria above.

* Age and gender information is available for everyone. Other information is missing for some individuals. Salary and hire information is based on 82,116, 52,788, 75,712, and 52,788 in each sample, respectively. Location (state) information is based on 68,534, 68,534, 48,854, and 48,854 in each sample, respectively.

^ There are 128 individuals whose age is 65+ but the exact age is unknown. We use 70 for these individuals to calculate the summary statistics.

Table 2: Year-to-year pass rates

BMI	Pass in year t	Fail in year t
Pass in year t-1	35,360 0.94	2,236 0.06
Fail in year t-1	1,983 0.18	8,777 0.82

Blood Pressure	Pass in year t	Fail in year t
Pass in year t-1	37,972 0.91	3,849 0.09
Fail in year t-1	4,185 0.64	2,350 0.36

Cholesterol	Pass in year t	Fail in year t
Pass in year t-1 (actual standards)	19,229 0.91	1,927 0.09
Fail in year t-1 (actual standards)	6,141 0.59	4,287 0.41
Pass in year t-1 (2011 standards)	22,034 0.90	2,317 0.10
Fail in year t-1 (2011 standards)	3,336 0.46	3,897 0.54

Glucose	Pass in year t	Fail in year t
Pass in year t-1 (actual standards)	13,046 0.94	788 0.06
Fail in year t-1 (actual standards)	15,096 0.85	2,654 0.15
Pass in year t-1 (2011 standards)	26,152 0.94	1,571 0.06
Fail in year t-1 (2011 standards)	1,990 0.52	1,871 0.48

Tables present observation counts in each cell (shares below the counts). An observation is an individual observed over two consecutive years. An individual who is observed over three or four consecutive years is included as 2 or 3, respectively, multiple observations. Cholesterol and Glucose measurement sarted in 2010 only, and passing standards were relaxed in 2011, so we report above passing rates based on either actual standard or the relaxed 2011 standards.

Table 3: Participation rates for eligible individuals

Number of years participated		Number of years eligible: 1	2	3	4	Total
	0	1,038 0.44	649 0.33	1,012 0.24	1,668 0.13	4,367 0.20
	1	1,347 0.56	420 0.22	478 0.11	1,241 0.09	3,486 0.16
	2		877 0.45	619 0.15	1,291 0.10	2,787 0.13
	3			2,157 0.51	1,978 0.15	4,135 0.19
	4				6,989 0.53	6,989 0.32
	Total	2,385 1.00	1,946 1.00	4,266 1.00	13,167 1.00	21,764 1.00

Table presents the number of individuals in each cell (shares below the counts). Sample restricted to individuals in the "complete data" sample who were eligible for participation in 2012.

Table 4: Comparison across short-run and long-run program exposure

	Dep. Variable: Pass in ...			
	BMI	BP	Choles.	Glucose
	(1)	(2)	(3)	(4)
Males				
Four-year participation	0.051	0.038	0.111	0.069
	(0.018)	(0.017)	(0.020)	(0.013)
Mean of dependent variable	0.80	0.84	0.70	0.60
R squared	0.008	0.016	0.136	0.681
N (individual-years)	13,253	13,253	10,091	10,091
N (unique individuals)	3,767	3,767	3,767	3,767
Females				
Four-year participation	0.042	0.046	0.034	0.031
	(0.017)	(0.012)	(0.017)	(0.009)
Mean of dependent variable	0.78	0.91	0.76	0.63
R squared	0.006	0.026	0.057	0.788
N (individual-years)	16,050	16,050	12,223	12,223
N (unique individuals)	4,569	4,569	4,569	4,569

Sample restricted to the 1,347 individuals who were eligible in 2012 and participated in the program that year, and the 6,989 individuals who were eligible for all four years (2009-2012) and participated in all for four years. Cholesterol and Glucose measurement sarted in 2010 only. In adddition to the reported coefficient, each regression includes year fixed effects and individual age fixed effects. Standard errors in parantheses.

Table 5: The relationship between health measures changes and healthcare expenditure

	Dep. Variable: Log(medical expenditure + 1)				
BMI	0.0174				
	(0.0074)				
BP (Systolic)		0.0036			
		(0.0012)			
BP (Diastolic)			0.0043		
			(0.0017)		
Cholesterol				0.0010	
				(0.0007)	
Glucose					-0.0003
					(0.0014)
Mean of dependent variable	6.20	6.21	6.21	6.23	6.22
N (individual-years)	55,373	55,974	55,909	39,234	39,232
N (unique individuals)	23,562	23,737	23,718	19,667	19,650

Table reports coefficients and standard errors (in parentheses) from regressing log(annual medical expenditure+1) on each biometric measure separately. Observations are individual-years in the "complete data" sample restricted to individuals who got their biometrics measured. Biometric measurements are missing for some individuals who auto-passed by passing all five metrics by satisfying NIH standards in the previous year or received exemptions. For each biometric, we dropped observations below the 0.5th percentile and above 99.5th percentiles. Cholesterol and Glucose measurement sarted in 2010 only. In adddition to the reported coefficient, each regression includes individual fixed effects and year fixed effects. Standard errors are clustered at the individual level.

Appendix Figure A1: Relationship between BMI and health expenditure

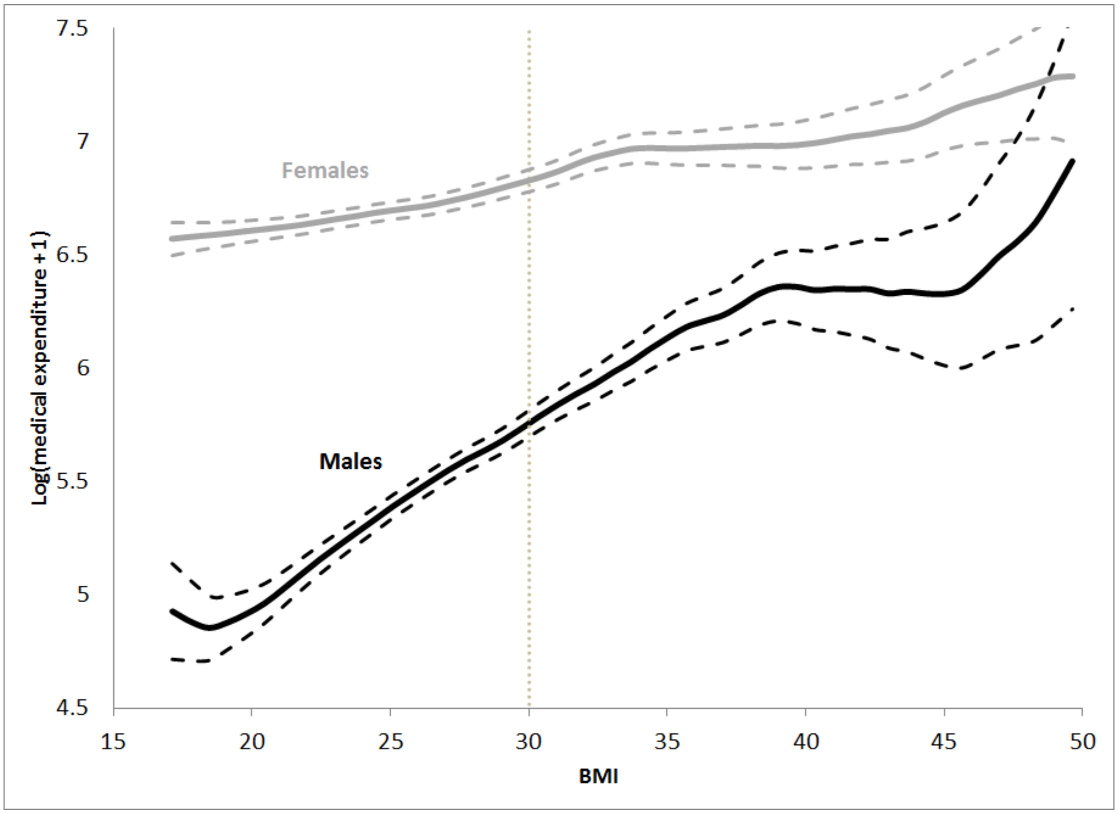

Figure plots predictions from kernel regression of log(yearly medical expenditure + 1) on BMI, separately for men and women. The dashed lines show the 95 percent confidence intervals and the vertical dotted line show the employer's passing standard (of 30). Observations are individual-years in the "complete data" sample for individuals who got their BMI measured. BMI measurements are missing for some individuals who auto-passed by passing all five metrics by satisfying NIH standards in the previous year or received an exemption (e.g., due to pregnancy). We dropped observations with BMI below the 0.5th percentile and above 99.5th percentiles. The sample includes 55,373 individual-year observations, consisting of 10,892 unique men individuals and 12,670 unique women.

Appendix Table A1: Passing standards in the Health Measure program, 2009-2012

Program Year	BMI	BMI Alternative (waist circumference)	Blood Pressure	Cholesterol	Glucose	Nicotine
2009	< 30	< 40 inches (men) < 35 inches (women)	< 140/90 mmHg	--	--	Negative Result
2010	< 30	< 40 inches (men) < 35 inches (women)	< 140/90 mmHg	HDL > 40 mg/dL & LDL < 130 mg/dL & Triglyceride < 200 mg/dL*	< 116 mg/dL*	Negative Result
2011	< 30	< 40 inches (men) < 35 inches (women)	< 140/90 mmHg	Total Cholesterol^ < 220 mg/dL	< 116 mg/dL	Negative Result
2012	< 30	< 40 inches (men) < 35 inches (women)	< 140/90 mmHg	Total Cholesterol^ < 220 mg/dL	< 116 mg/dL	Negative Result

Table shows passing standards for each health measure. BMI is the ratio of the individual's weight (in kilograms) to the square of her height (in meters). Alternatively, individuals can pass the BMI metric with waist circumference measure. Blood pressure is measured using systolic and diastolic mmHg readings. In 2009, cholesterol testing was only required of those in a "high risk" demographic group. Those at "low risk" received the cholesterol incentive by filling out a documentation form. Cholesterol and Glucose measurement started in 2010 only, and passing standards were relaxed in 2011.

* In 2010 individuals had to pass both cholsterol and glucose to receive incentives associated with either of the two.
^ Total cholsterol is equal to LDL + HDL + triglycerides/5.

Appendix Table A2: Financial incentives associated with passing

	BMI	BP	Choles.	Glucose	Nicotine
PPO enrollees in Choice Fund I					
2009	6	1.5	--	--	6
2010	6	4	4	4	6
2011	4	4	4	4	4
2012	5.5	3	3	3	5.5
PPO enrollees in Choice Fund II					
2009	3	0.75	--	--	3
2010	5	2.5	2.5	2.5	5
2011	3	3	3	3	3
2012	4.5	2	2	2	4.5

Table shows incentive amounts (in current $US, per week). The first panel shows the weekly incentives received (if passed) by individuals enrolled in "Choice Fund I" (the higher coverage PPO health plan) and second panel shows the weekly incentives received by individuals enrolled in "Choice Fund II" (the lower coverage PPO health plan). We do not observe which PPO enrolees are in, but our understanding is that the majority of individuals were covered by Choice Fund I.