**Step 1**: Count the number of claims by employee

I would begin by counting the number of claims submitted by each patient (using some sort of de-identified member ID number) to see the distribution of the number of claims. I would create a histogram to visualize the frequency distribution which would presumably be right-skewed. That is, a large portion of the population would have five or fewer medical claim events while a long tail of very few employees would stretch into the 20s or maybe even 30s.

**Step 2:** Model the relationship between number of claims and reimbursement amounts

Logic dictates that as the number of claims increases so too does the plan payouts. I would run a regression to test this assumption and create a scatter plot to visualize the relationship. While this seems obvious we need to confirm it as we work toward our objective of identifying a condition to target. It’s even possible that two have a nonlinear relationship once claims number surpass a certain point. Thus we might want to consider a log transformation of our independent variable(s) to better capture any nonlinearity.

**Step** 3: Identify health care “super users”

The pareto principle states generally that 80% of the effects come from 20% of the causes. The health care corollary then is that 20% of users, we’ll call them our super users, are responsible for 80% of the plan costs. We can identify them by binning our data (in Excel one would use *=percentile*, in Pandas the *.quantile* method) at the cut-off point of our choosing. I would take a bullish approach and say that those in 99th percentile are our super users. That is, only those with an extremely high number of claims (probably somewhere in the high teens or above) would get classified thusly. We could then create an *IsSuperUser* column with an indicator flag (1 for yes, 0 for no) for each employee.

**Step 4**: Identify health condition(s) most associated with *IsSuperUser* flag

There are two ways to accomplish this. First, create a correlation matrix to see how each variable (column) is correlated with *IsSuperUser*. Presumably, some chronic conditions will be positively correlated with this status, meaning that they move in tandem. Another way is to create 2x2 pivot tables showing how the count of claims are distributed for different chronic conditions among super and regular users (example below).

|  |  |  |  |
| --- | --- | --- | --- |
| *IsSuperUtilizer* | N | Y | All |
| Has\_Alzheimers |  |  |  |
| Y | 21,800 | 610 | 22,410 |
| N | 93,418 | 524 | 93,942 |
| All | 115,218 | 1,134 | 116,352 |

This contingency table shows us, among other things, that we have 1,134 super users in our employee population of ~116,000. It’s unclear (from an objective standpoint) how concerned we should be that ~53% of super users have Alzheimer’s. We need to compare these values with other conditions.

The condition below, ischemic (coronary) heart disease, is a much better candidate for intervention. Nearly 90% of those with the condition are super users (1,017/1,134). This means that if we could address the causes of the disease, it’s likely that we’d see fewer health claims as some of these super users become regular users.

|  |  |  |  |
| --- | --- | --- | --- |
| *IsSuperUtilizer* | N | Y | All |
| Has\_IschemicHeartDisease |  |  |  |
| Y | 47,925 | 1,017 | 48,492 |
| N | 67,293 | 117 | 67,410 |
| All | 115,218 | 1,134 | 116,352 |

**Next Steps:** Incorporate Rx event data

Of course, this is just one approach and does not even make use of pharmacy data which could offer a wealth of additional insights. Some ways we could incorporate the Rx data include:

* Count and graph the distribution of pharmacy events; aggregate and map the totals to each employee ID. This would tell us how many average pharmacy visits an employees has in a year and lets us identify high-usage employees.
* Sum or average total pharmacy spend by chronic condition. This could reveal whether certain chronic conditions are more “expensive” than others from a drug standpoint. For instance, what does an average employee with Alzheimers spend at the pharmacy vs an average employee with COPD? This is another good approach for identifying a condition for intervention.
* After incorporating Rx data, see whether these new variables are good predictors of super user status by running a logistic regression.
* Begin creating predictive models to determine which employees might become super users.

**Bonus**: Sample Analysis

Knowing that some of the descriptions above can be vague, I’ve performed most of the analysis outlined on this page using de-identified Medicare data from 2008 provided by cms.gov. [Analysis available on GitHub.com](https://github.com/camoser19/Healthcare-Analytics-Case-Study/blob/master/MorganStanley_CaseStudy_HealthcareAnalyticsManager_CoreyMoser_May2019.ipynb).

**Chronic Condition Selection**

My previous analysis has revealed that most health care super users in our population have coronary (ischemic) heart disease; if we could shift some of these users away from the doctor’s office and ER, and thus reduce their claim amounts and total spend, we have the potential for huge savings.

Coronary heart disease is not strictly “curable”, but it is treatable. Evidence suggests that normalizing blood pressure and lowering cholesterol can reverse plaque accumulation in coronary arteries. By addressing these risk factors patients are less likely to suffer from health symptoms that precede a heart attack (at which point it’s too late to do anything) such as chest pain; nausea; vomiting; and upper abdomen pain, discomfort pressure or tightness (to name a few).

**Vendor Background and Selection**

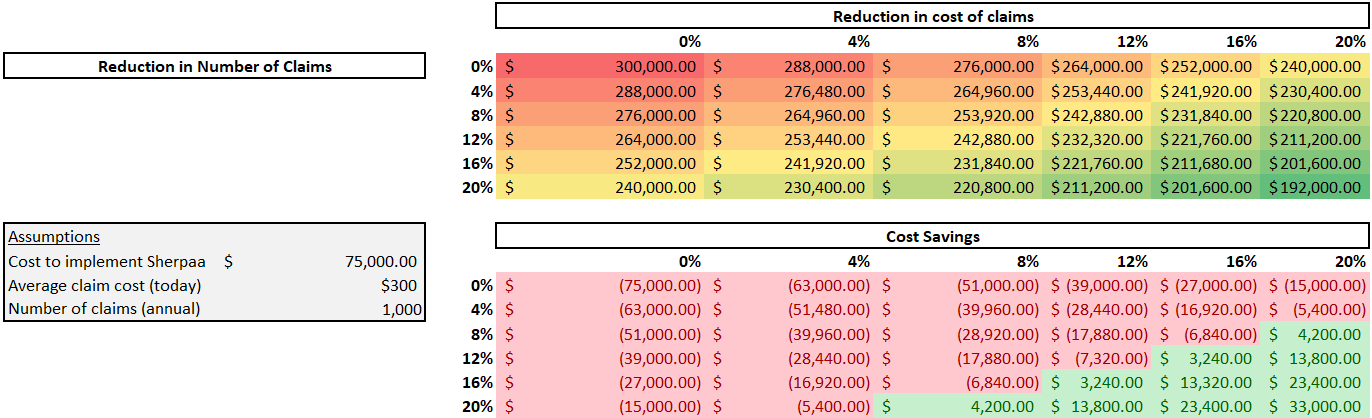
After some research I believe [**Sherpaa**](https://sherpaa.com/treats/you), a seven-year-old Brooklyn-based healthcare company, can help us accomplish this task. Sherpaa connects employees directly with doctors and insurance guides online to reduce healthcare costs. Their value proposition is simple: in-person doctors’ visits are too expensive and time consuming to treat the majority of health issues. They’ve identified 1,500 conditions, including imaging and tests, Rx refills and chronic conditions etc., that can be addressed with the same level of quality and positive health outcomes as in-person doctor visit. Employees can interact with doctors online, via their phone or desktop or by sending photos, before seeking in-person care, at a lower cost for employees and employers.

**Why Sherpaa?**

Sherpaa has experience treating and creating positive health outcomes around some of the abovementioned symptoms caused by coronary heart disease. For instance, they can work with employees and prescribe provide high cholesterol or heartburn medication to manage the discomfort that might otherwise cause employees to seek medical care. They can also coordinate lab and imaging cholesterol tests to help identify risk factors for developing coronary heart disease and recommend ways to address issues before they become chronic and require expensive intervention (such as surgery to replace a heart valve). In short, Sherpaa is geared toward addressing this prevalent population issue.

**Analysis**

The most important analysis is to determine return on investment (ROI). Will the health savings gained (in terms of fewer and/or less expensive claims) be greater than the cost of implementation? To determine this we could perform a sensitivity analysis (see example below):



This can help us determine the reduction, in either claim cost or number of claims, for Sherpa to realize for it to be worth the cost of implementation. This of course doesn’t account for time and effort from Morgan Stanley’s employees which would further influence our decision.

And while cost considerations are paramount it’s also important to measure quality of the service and employee experience. Will our employees use this service? Does it offer a clear value propositions that our employees will recognize? Do its marketing materials reflect its value or will it require effort from us to clearly communicate its benefits? What does Sherpaa’s technology infrastructure look like and how much effort will be required for it to pass a security architecture review? These type of questions cannot be overlooked.

Sherpaa claims 150+ clients including companies like Blue Apron, Betterment and Casper. I would reach out to these companies to find out how they like working with Sherpaa, whether they plan to continue using it, how their employees value the service and what usage rates look like.

I would ask for client testimonials (as well as feedback from users of the service within those companies) as well as data (where possible) showing how Sherpaa intervened to change health outcomes in other, similar employee populations. This last point is crucial, if I cannot understand how Sherpaa measures health outcomes I won’t be able to explain their value to our organization and thus lack leadership buy in.

**What I’d Ask of Sherpaa**

As with any vendor my first concern would be to ensure that they meet their contract guarantees. If Sherpaa says they can lower CHD-related spend or number of claims I’d make sure that was in their contract and, in case of breach, absolve Morgan Stanley of payment. Alternatively, I’d create an arrangement where payment could be made quarterly as certain targets are reached.

Other things I’d want to see include:

* Sign up rates: What percentage of client populations enroll and what can we expect if we follow a recommended marketing cadence?
* Usage rates: How many interventions do you expect to make per month? Per quarter? How many monthly active users should we expect?
* Number of interventions: Sherpaa claims to, “prevent 70% of office, ER, and urgent care visits from happening.” How is this determined? Can you prove you were or are doing this for our population?
* Reduction in CHD-related claims and Rx fills: In tandem with Sherpaa I would identify a number of prescription drugs and claim types that could reliably sourced to CHD-type illnesses. Drugs to treat cholesterol and blood pressure could qualify. So would any heart surgery or other lab or imaging procedures related to heart health. Once agreed upon I would expect to see Sherpaa reduce these numbers.

**What I’d Do Internally**

* Review email open and clickthrough rates to see how employees responded, generally, to Sherpaa’s launch.
* Conduct satisfaction surveys among users and general awareness surveys among the wider employee population.

**Other Analyses**

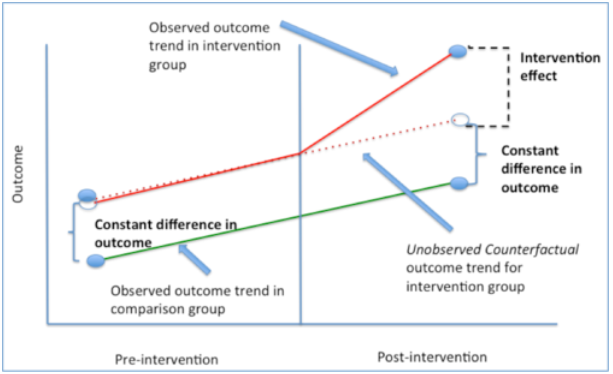
* Among super users, what percent of their claims and Rx spend is attributable to CHD-related treatments? I would want to see this number go down over time.
* Similarly, I’d hope to see fewer super users overall. If we could eliminate some of their CHD-related claims they could fall into the merely “heavy user” category. This would be one of the biggest measures of success in my opinion.

Figure 1: Graphical representation of difference-in-difference technique.  
Source: <https://www.mailman.columbia.edu/research/population-health-methods/difference-difference-estimation>

* Another analysis I’d be interested in is a differences in differences estimation. This would work especially well with a phased roll out if we make this benefit available only to a portion our employee population. Essentially, we identify two populations and construct them to be as similar as possible. One is our control group (who will not receive the Sherpaa benefit) and the other is the treatment group (which does). With proper historical data we should be able to predict our CHD spend in the future. This is our “trend” and should be the same for our both groups (assuming we’ve constructed them properly). After a year with Sherpaa we should see whether the treatment group has a lower healthcare spend than what the trend would otherwise predict. If we don’t see the trend change for the treatment group, then it’s possible that the program isn’t working as promised.
* Are we seeing positive spillover effects in any other areas we can measure? General improvements in population health can cause a ripple effect that can generates financial returns elsewhere in the form of shorter inpatient stays and lower readmissions rates. Without at least trying to quantify some of these effects we may be missing some of the benefits not captured by a strict ROI measure.