



How to tell if your vendor's claims are valid: Part One

Inaccurate marketing claims and outcomes reports are proliferating. The Validation Institute has staked out a position as the leader in assisting/promoting vendors and consultants in the "Integrity Segment" of the healthcare services market.

How can you tell if your adviser is in the Integrity Segment? The easiest way: did they send you to this series or did you have to find it on your own?

.....

Part One will cover regression to the mean (RTM). That is the tendency of outliers on the high or low side, over time, to come closer to the average. For instance, every season several baseball teams will win or lose their first 4 or 5 games. Yet over the course of a full season, no team has ever won more or lost more than 75% of its games. And over the history of major league baseball, no team has won or lost more than 58% of its total lifetime games.

What makes employee healthcare services RTM different from that baseball example is that while it is very easy to identify and "count" outliers on the high side who decline in risk or cost, only a handful of consultants and buyers consistently also measure outliers on the low side. As you will see at the end, a small number of vendors will also measure this way. The Validation Institute welcomes vendors who are willing to do that. We also welcome vendors who genuinely believed they were measuring validly, but after reading this article realize that they are inadvertently overstating impact. And as part of your Employer RFP Support package, you have received or will receive an RFP that elegantly adjusts for RTM no matter how much the vendor tries to embrace it.

Otherwise, this one-way RTM creates a built-in bias towards showing savings where none exist. Let us first examine this in theory, and then we will present examples of vendors doing this in practice.

Suppose your company, Vandalay Industries, has two asthmatic employees, George and Jerry. A vendor/carrier proposes an asthma program. George has an asthma attack generating a \$2000 ER visit in the Baseline year. Jerry doesn't, because they don't know he is asthmatic. George qualifies for the program. During the contract year Jerry has a \$2000 asthma attack. George doesn't.



| | Baseline year | Contract Year |
|---------------|---------------|---------------|
| George | \$2000 | \$0 |
| Jerry | \$0 | \$2000 |
| Average spend | \$1000 | \$1000 |

Obviously, no asthma attacks were avoided and Vandalay Industries saved no money. But that’s not the way vendor math works.

Here’s why. The vendor doesn’t measure Jerry because Jerry didn’t have any asthma-related claims even though he has asthma. Non-validated vendors either don’t know -- or know but won’t admit -- that people like Jerry who need help but don’t have claims even exist.

Therefore, the vendor only reports on the observed, high-cost George...and inadvertently or deliberately fabricates improved outcomes and savings for the “population” as a whole -- meaning just George.

| | Baseline year | Contract Year |
|---|---------------------|---------------------|
| George | \$2000 (one attack) | \$0 |
| Jerry | \$0 | \$2000 (one attack) |
| Average number of asthma attacks | 1 | 1 |
| Average spend | \$1000 | \$1000 |
| Vendor-reported number of asthma attacks in the entire asthma “population.” | 1 | 0 |
| Vendor-reported claims spend in the entire asthma “population.” | \$2000 | \$0 |

Now consider a similar chart, but over 4 years. Instead of asthma, this is smoking. Instead of disease management, this is a wellness vendor’s smoking cessation program. And instead of asthma attacks, George and Jerry routinely quit smoking and then start smoking again, like many smokers do. In this example, assume they quit/resume every other year. Also assume that to qualify for a smoking cessation program, you must be a current smoker, of course.

| | Baseline Year | Contract Year 1 | Contract Year 2 | Contract Year 3 |
|--|---------------|-----------------|-----------------|-----------------|
| George | Smokes | Quits | Smokes | Quits |
| Jerry | Quits | Smokes | Quits | Smokes |
| Average number of smokers | 1 | 1 | 1 | 1 |
| Reduction in the number of smokers vs. baseline | 0% | 0% | 0% | 0% |
| Reduction in smokers as reported by the vendor | 100% | 100% | 100% | 100% |

A vendor can keep reporting 100% reductions in smoking every year, even as the number of smokers remains unchanged!

Not just a curious anomaly, this statistical sleight-of-hand has helped to fuel the rise of the wellness, diabetes, and other “point solutions” industry.

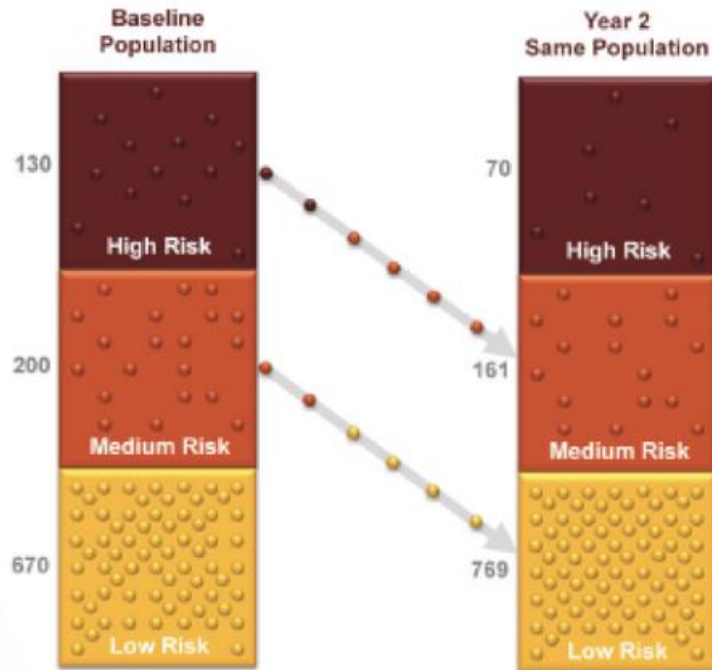
Let’s consider several examples of vendors and carriers doing this in practice.

So far in this article, we have been careful to acknowledge that some genuinely honest vendors, consultants and actuaries simply don’t understand this fallacy, not having taken our course. The vendors cited below, on the other hand, are quite aware of what they are doing, as is obvious from their own displays, data and/or statements.

Here is an example of a vendor that “guarantees” a reduction in high-risk members. They don’t say anything about low-risk members increasing in risk.

Oversight? We think not.

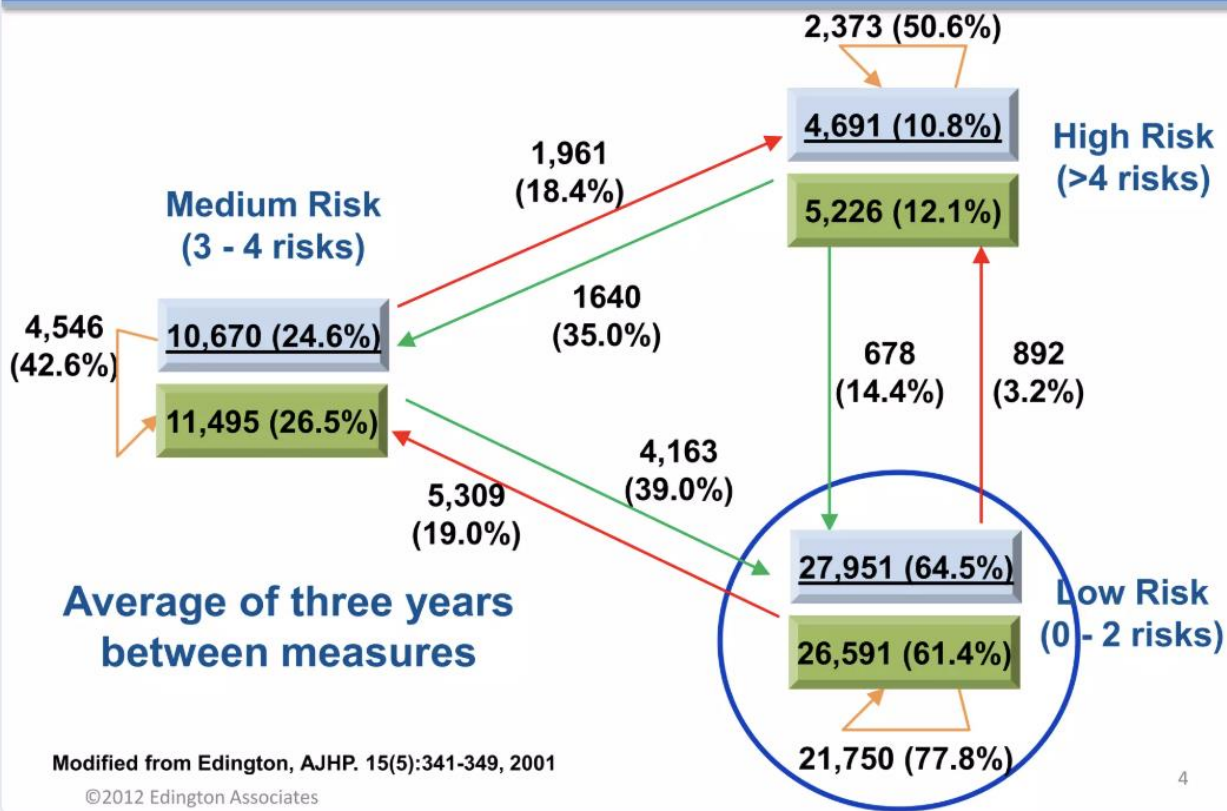
If you examine their “bar chart” carefully, you’ll see that they’ve deliberately put the two-thirds of employees who are low-risk into the lowest one-third of the graph, in order not to draw attention to the fact that the large majority of employees are low-risk. In case you are wondering how they justify this sleight-of-hand arithmetically, count the bubbles. Each bubble represents 10 people.



This is exactly the same fallacy as above, with more realistic assumptions. Just like smokers quit and resume, people gain and lose weight. Blood pressure can go up or down. Bottom line: people do change risk levels naturally, wellness program or not.

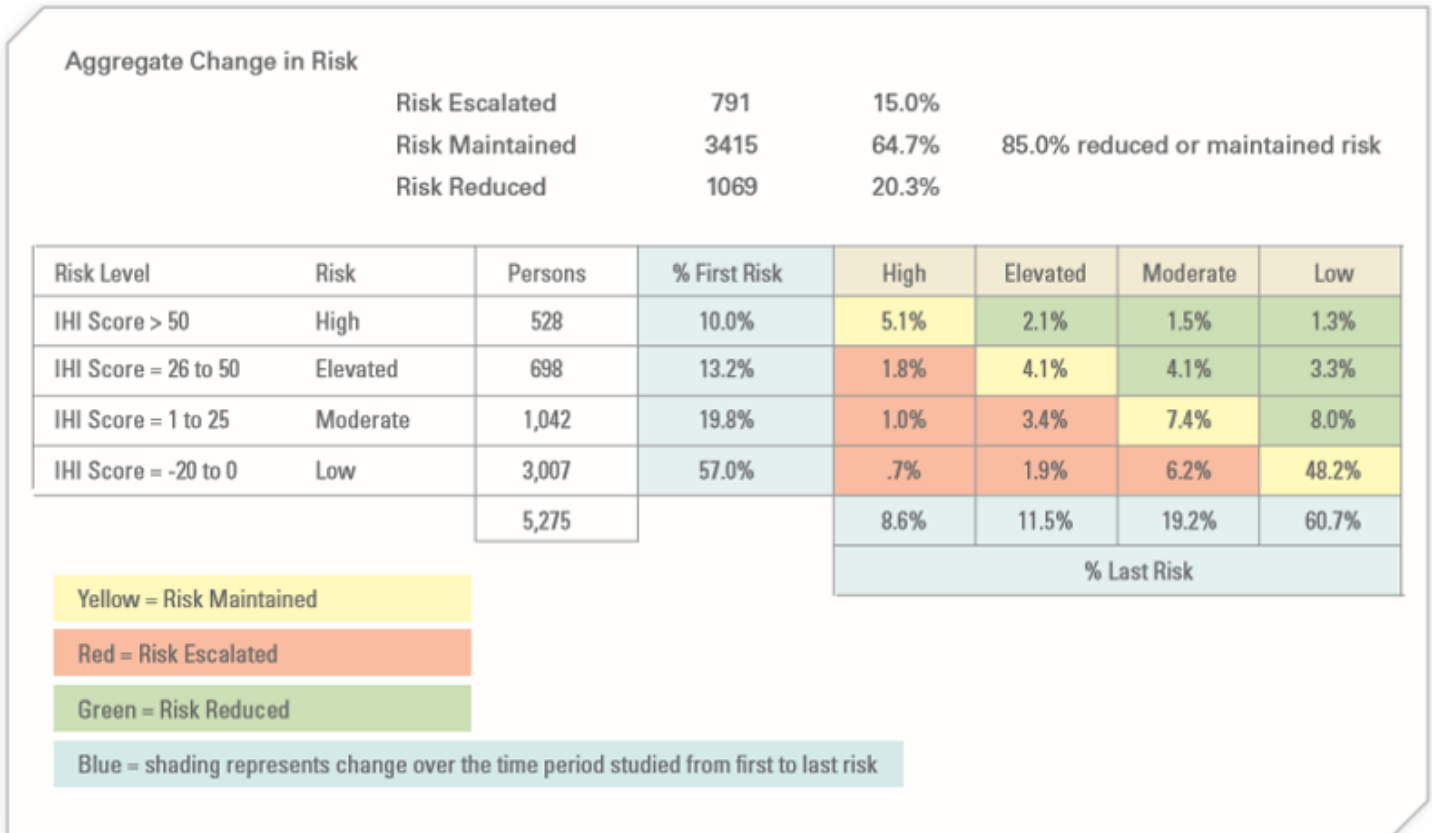
Take a few minutes (and it will take a few minutes) to digest this [“natural flow of risk” chart, developed by Dee Edington](#). (Scroll to Slide #4.)[1] As you can see, absent any intervention, people do fairly predictably change.

Natural Flow: by Risk Status



[1] The slide alone can be found here.

Here is how Interactive Health presented this fallacy before they went bankrupt. They divided the population into four risk categories and showed how people migrated between the two periods from one risk category to another.



Green improved between the two periods, yellow stayed the same, and red got worse. The change between the two periods was mostly regression to the mean, according to their own chart.

You might say: “Well, at least they were being honest.” Haha, good one. Obviously, you are new to the industry. No, here’s what they claimed this chart showed. They said 49% of high-risk members improved, while only 8.8% of low-risk members got worse. True enough. Of the 10% high-risk to begin with, 4.9% improved and only 5.1% stayed the same. Likewise, their percentages are correct for low-risk members getting worse.

But now consider what really matters—the *number* of employees, not the percentage. Because very few were high-risk to begin with, that 49% who improved constituted 259 employees. On the other hand, so many employees -- 3007 -- were low-risk to begin with that the 8.8% who got worse equaled 264 employees. That’s five more than the number who improved. You would never guess it from those percentages they presented.

The heading is that “85.0% reduced or maintained risk.” But another way of reading that exact same summary data would be that: “79.7% failed to improve their risk scores.”

The Ultimate Perp...Apprehended!

One vendor, Wellsteps, actually admitted to milking regression to the mean, in order to fabricate savings. As a result, they and their customer, Boise Schools, won the wellness industry's highest honor, the C. Everett Koop Award. This was predicted because the award is generally given to people connected to the awards committee (their CEO had been on it) whose programs are especially fallacious.

Their many fabricated claims included this one:

A mixed model repeated measures evaluation³ was also applied to employees who demonstrated the worst health behaviors at baseline. Table 3 shows the program impact on individuals who had the worst health behaviors at baseline. After 12 and 24 months, dramatic improvements were seen in those employees who had previously reported low levels of baseline exercise, fruit and vegetable consumption, days of restful sleep, and high levels of tobacco and alcohol use.

Table 3. Improvements in Health Behaviors Among Those with the Worst Health Behaviors³

| Health Behavior | Baseline | Year 1 | Year 2* |
|---|---------------|---------------|----------------------|
| <2 days of exercise/week, n=502 | 1.46 days | 2.29 days | 2.47 days (+69%) |
| <60 minutes/week, n=373 | 18.79 min | 113.20 min | 134.71 min (+617%) |
| <3 daily servings fruits/veggies, n=426 | 2.46 servings | 3.48 servings | 3.63 servings (+47%) |
| <3 days of restful sleep/week, n=407 | 2.23 days | 3.17 days | 3.36 days (+51%) |
| Smokers (days per week, n=77)** | 4.35 days | 5.43 days | 4.27 days (-1.6%) |
| Alcohol Use (drinks/day, n=691) | 1.31 drinks | 1.16 drinks | 1.10 drinks (-16%) |

*All changes are statistically significant (p<.05)

** 11 of 77 smokers quit at 2 years

In other words, they cherry-picked the worst employees and showed an improvement in them. Along with the unaddressed curiosity factor of how they conclude that chain-smokers only smoke 4 or 5 days a week, this is as pure a regression-to-the-mean play as exists anywhere. [Their CEO, Steve Aldana, did admit that, to STATNews, as part of an exposé on their program:](#)

“In just one year, many employees will move from one [risk] group to the other,” [Mr. Aldana] explained, “even though they did not participate in any wellness programs or any intervention whatsoever.”

That movement, he continued, “reflects changes in health risks that occur naturally,” making it possible that some high-risk people become low risk “even though your program didn’t do anything.”

This program may have fooled the Koop Award committee, but it didn’t fool STATNews. Nor did it fool the Boston Globe, which [devoted a full-page spread](#) to Wellsteps actually making Boise School employees worse while pretending to improve them:

Top wellness award goes to workplace where many health measures GOT WORSE

The average BMI of the 800 people who started the program was 27.5, according to the study. That of the 700 employees who did not participate was 28.5. The average BMI of the 800 who did not participate was 28.5.

Wellness programs are not fooling anyone, says a new study. The average BMI of the 800 people who started the program was 27.5, according to the study. That of the 700 employees who did not participate was 28.5. The average BMI of the 800 who did not participate was 28.5.

Wellness programs are not fooling anyone, says a new study. The average BMI of the 800 people who started the program was 27.5, according to the study. That of the 700 employees who did not participate was 28.5. The average BMI of the 800 who did not participate was 28.5.

And now it shouldn't fool you.

The Take-Aways

If you as an employer are reading this, you also have access to our Valid Procurement



First, any vendor whose pitch includes: “We manage your highest...” should immediately be told to go get validated by the Validation Institute (VI) for that claim and then come back later. The honest ones will follow up. The others? This will be the last you hear of them.

If a vendor includes anything like this in their outcomes report, demand a refund. Those vendors are fully aware of this fallacy.

Second, dismiss any claim that compares percentage of improvers to percentages that get worse. It’s all about the absolute numbers.

Third, every VI-validated vendor offers a metric that controls for regression to the mean. (Attention, validated vendors: this isn’t to say this control is required as part of your validation. However, it is required to be offered. Many employers won’t have the data to take you up on the offer.)

For instance, US Preventive Medicine specifically measures its risk reduction against that which would happen naturally. Virta is the only diabetes vendor that will allow you to contract for a simple comparison of the number of uncontrolled diabetics year over year across your population. That comparison avoids both regression bias [and participation bias](#). The latter is another huge fallacy to be covered in our [next installment](#).

Finally, please enroll in both our [Critical Outcomes Report Analysis \(CORA and CORA Pro\)](#) courses, which cover this and other fallacies. It is especially important to urge your consultants and brokers to enroll, as you are paying them to protect your interests, but the proliferation – even the continued existence -- of these fallacies strongly suggests they aren’t.

[COMING NEXT](#)

Participants-vs-non-participants: the latter does not control for the former. Not even close.