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

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?

Now, it is time to put everything together to write the RFP and the contract. Besides the usual litany of questions that consultants include even though the answers are pretty tepid, you should ask the questions in Part Five.

Ask the Validation Institute to recommend brokers and consultants who adhere to these Validation Institute principles for writing RFPs. You'll find that using and advisor to ask the right questions rather than the most questions is both more effective and more affordable.

Before we delve into the substance of making the outcomes valid and comparable, there is one key question to ask first if you are getting pitched by a carrier, and one if you are getting pitched by a vendor.

Key Carrier Question

This one applies only to solutions that your carrier or broker is pitching, as opposed to those that you seek on your own, or from vendors. However, in those situations, it is probably the single most important "bright line" question. If you were only allowed to ask one question, this would be the one:

> *"What is the penetration of this solution in your own insured (or in the case of large consulting firms, covered) population?"*

Many carriers will resell solutions that they themselves don't use, simply because there is a (typically undisclosed) markup. Or if they offer access to fully insured members, a fully insured member really must know how to find and insist upon them. If it's not good enough for the carrier to actively promote to its own employees or members, why should it be good enough for you?

There is one exception to this rule: carriers often quote their fully insured customer an initial price, and then offer to reduce the price if the customer agrees to use their wellness program. Of course, the sum total of the lower price plus the wellness program price will always exceed the initial price, because wellness doesn't come close to paying for itself. If it did, the customer would pay a lower price for both in combination than just for insurance alone.

There is even an exception to this exception: one carrier deliberately sets a high price for fully insured customers, so that the price with wellness is lower than the price without. In that case, as a fully insured customer, demand the lower price, with a $100 PMPY subtraction -- and pass on the wellness program. Say: "Sure, we'll happily add back the $100 and take on the wellness program -- as soon as its outcomes are validated by the Validation Institute."

Excellent examples of add-ons that fully insured groups don't pay extra for would be utilization management and case management. Clearly, the carriers know these services pay for themselves. Likewise, they know when these services don't pay for themselves.

Key Vendor Question

All the fallacies in the previous six installments can be "controlled for" with a single question to a vendor:

> *"If we promote your solution to half our population and not the other half (which will have access, but not promotion), how much of your fees will you put at risk that the first half will outperform the second half in the key metrics you are addressing?"*

Unless there is a bright-line distinction between subgroups within your population (such as a union contract), you can't discriminate in your benefits. You can, however, discriminate in your promotion of those benefits, provided the Summary Plan Document is the same for all like-positioned employees.

A really good vendor will offer to put 50% to 100% at risk (if both groups are large enough) and say: "Where do I sign?"

There are a couple of asterisks. First, you'd have to trim outliers (unless the program itself is for outliers). Second, you'd need a couple of years of history to see if there is a difference in the groups to begin with. Because you are looking at a difference as opposed to an absolute number, 2020 can be a baseline.

Making the outcomes comparable and valid

The classic mistake in an RFP is to ask the respondents what they will achieve instead of telling them precisely what you expect them to achieve. Asking "What's your ROI" and/or "How do you measure ROI?" immediately shows the vendors they can say whatever they like.

By contrast, If you lay out all the specifications for measurement right in the RFP, the responses are both valid and comparable to one another.

What happens if you don't make the rules can best be illustrated by an offer routinely made by an extremely major carrier (name on request, following an NDA) concerned that valid measurement would (quite justifiably) show no impact of their worthless health risk assessment. They offer three "choices" for how to measure outcomes. While most buyers and advisers acquiesce to one or more, those who have completed the Validation Institute's advanced course in analyzing outcomes, CORA Pro, see right through them.

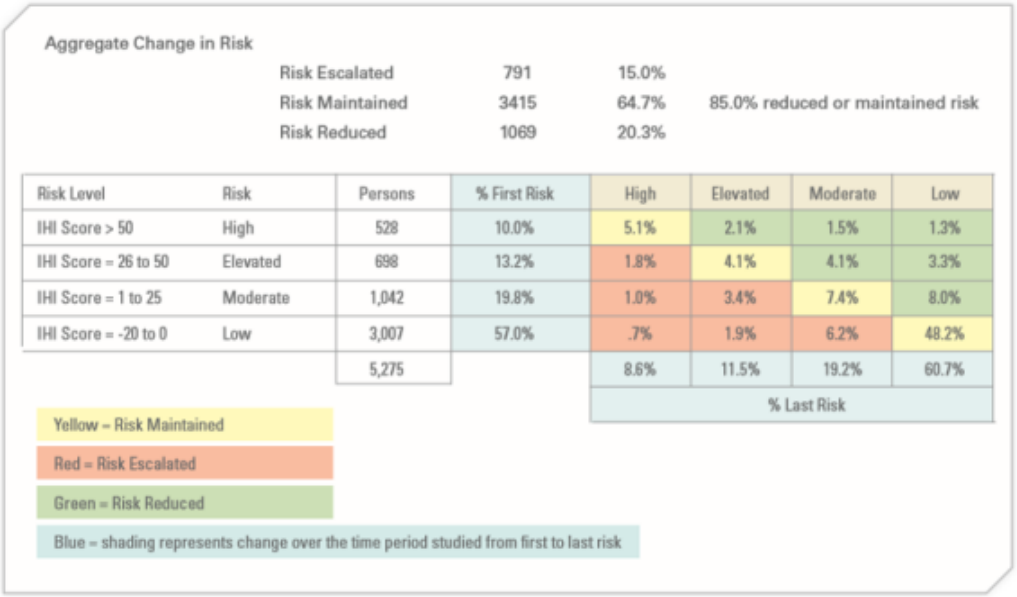
Test your knowledge on these three measurement options:

A. For any two adjacent reporting years, at least X% of those who were at medium or high risk move to a lower risk status (e.g., high to medium, high to moderate or low, or moderate to low). The value of X is negotiable.
B. Alternatively, net risk reduction percentages can be calculated, by considering movement to both higher or lower risk. In this instance, the percent of HRA completers who increase risk should be lower than the percent who decrease risk (subject to engagement in risk-reduction programs). The performance guarantee could be based upon how high the difference is between the people who improve risk vs those who do not.
C. Similar to a and b, but only for those who participate in wellness coaching.

Having read the first six installments, you should be able to identify the invalidity in each one of these. "A" is classic regression to the mean, covered in Part One. "C" adds pure participation bias. "C" was debunked in Part Two.

"B" is more subtle. A cursory reading suggests that both increases and decreases in risk are taken into account, removing the regression to the mean and hence being a valid metric.

Before acquiescing in this, revisit the visuals from Part One, particularly this one:



Aggregate Change in Risk

Risk Escalated	791	15.0%	
Risk Maintained	3415	64.7%	85.0% reduced or maintained risk
Risk Reduced	1069	20.3%	

Risk Level	Risk	Persons	% First Risk	High	Elevated	Moderate	Low
IHI Score > 50	High	528	10.0%	5.1%	2.1%	1.5%	1.3%
IHI Score = 26 to 50	Elevated	698	13.2%	1.8%	4.1%	4.1%	3.3%
IHI Score = 1 to 25	Moderate	1,042	19.8%	1.0%	3.4%	7.4%	8.0%
IHI Score = -20 to 0	Low	3,007	57.0%	.7%	1.9%	6.2%	48.2%
		5,275		8.6%	11.5%	19.2%	60.7%
				% Last Risk			

Yellow = Risk Maintained
Red = Risk Escalated
Green = Risk Reduced
Blue = shading represents change over the time period studied from first to last risk

Watch what happens if you use the carrier's "recommendation," vs. what really happens in the population.

Assume for simplicity the only two categories being tallied above are the extremes: the high- and low-risk.49% of the high-risk cohort enjoyed a decline in risk, while only 8.8% (57% - 48.2%) of the lowest-risk category increased. The carrier hits their numbers and maybe earns a bonus, using their "b" or "c" recommendation. (Obviously on "a" they can't lose.)

Now, instead of looking at the percentages, as the carrier wants us to, look at the actual number of people whose risk factors migrated. You do this by multiplying those two percentages above by the number listed in the "persons" column. You'll find that only 258 of the 528 high-risk "persons" declined in risk, while 264 of the 3007 low-risk "persons" increased in risk. The carrier misses their numbers. Their sleight-of-hand of offering to tally per cents rather than raw numbers skews the result from completely valid to completely invalid.

Ergo, you must make the rules yourself. This ensures validity. Equally importantly, this allows an apples-to-apples comparison of RFP responses. We recommend that the key to the RFP be a table like this, with your range of goals across the top and the percent fees at risk along the side. These need not be the exact ones here. The important thing is to lay out specifications, rather than ask for an open-ended answer.

Percent of fees at risk	Reduction in [fill in the blank] by Year 1,2 or 3		
	3%	6%	9%
20%	Price	Price	Price
40%	Price	Price	Price
60%	Price	Price	Price
80%	Price	Price	Price

The fill-in-the-blank depends on what you are RFP-ing, of course. It could be a decline in risk factors, to use the example above. Or, if a diabetes program, a decline in the total number of people whose Hb A1c's exceed 9.0. it could be a decline in spinal fusions, for a musculoskeletal RFP. Or, for a prenatal RFP, a decline in neonatal days. In all cases, it would be across the population. Not just participants or high-risk "persons" or a study group compared to "matched controls."

The price quoted will vary according to the risk taken and the goal. It would not be surprising if highest risk and biggest stretch outcomes were priced 50% higher than the low-risk, low-return outcome.

This would be particularly true if you have high turnover, since many of these programs take three years to show any noticeable impact, and even then it is still a crapshoot. You will need to disclose your turnover rate in the RFP as well.

Vendors need not complete the entire table. Wellness and diabetes vendors are usually only willing to put their profit at risk (20%), rather than take an actual risk with their cost. By contrast, other vendors such as finHealth (claims auditing), Quizzify (employee health literacy) and Sera Prognostics (prenatal) will go to 60% or even 100% fees at risk.

How much of the table a vendor completes is data by itself. If they don't think they can hit 9%, why should you?

Once completed, you can make valid comparisons at a glance, for the cells that are completed.

In some categories, like risk factors or prenatal, the denominator is self-evident. In others, you have to be very specific about the denominator. For instance, in wellness, diabetes, or disease management, you list the ICD-10s (or DRGs) where you expect to see the reduction. Collectively, these are known as "wellness-sensitive medical event rates," or WSMEs.

Here is a list of ICD10s that the Validation Institute recommends to capture those WSMEs:

Chronic Condition	***ICD10 Codes (includes all subcategories to the basic codes, even if they are not listed)***
Asthma	J45
Chronic Obstructive Pulmonary Disease	J40, J41, J42, J43, J44, J47, J68.4
Coronary Artery Disease (and related heart-health issues)	i20, i21, i22, i23, i24, i25.1, i25.5, i25.6, i25.7
Diabetes --including likely non-cardiac complications	E10, E11.0-E11.9, e16.1, e16.2, e08.42, e09.42, e10.42, e11.42, e13.42, e08.36, e09.36, e10.36, e11.311, e11.319, e11.329, e11.339, e11.349, e11.359, e11.36, e13.36, L03.031, L03.032, L03.039, L03.041, L03.042, L03.049, L03.115, L03.116, L03.125, L03.126, i96, E08.621, E08.622, E09.621, E09.622, E10.621, E10.622, E11.621, E11.622, E13.621, E13.622, L97
Hypertension, Heart Failure and related diseases	i50, i10, i11, i12, i13

Many vendors will balk at promising a reduction in a variable without knowing what that variable is. This is perfectly reasonable. What if you have a low-risk population to begin with? You probably shouldn't be doing wellness or diabetes if you do, of course.

That means you also need to give them baseline data. You would have your own risk factor data in-house, or you could do a baseline screening. On the other hand, claims data can be time-consuming to obtain from your carrier. It is your data, though, so you are entitled to it. And you need to know it anyway. Without knowing your baseline, how would you know how much of a reduction to expect, or whether it got achieved?

Nonetheless, obtaining this data will take weeks. If you are under a deadline, we recommend starting with Validation Institute averages, and then saying that the guarantee will be adjusted down if indeed the baselines are lower, by the same percentage. For instance, in the commercially insured population, these might be the numbers, per 1000. (The VI's actual numbers are proprietary.)

Category	Rate per 1000 of ER plus IP
Asthma	2.0
Coronary Artery Disease	1.6
CHF/hypertension	1.8
Chronic Obstructive Pulmonary Disease	1.4
Diabetes (including common complications)	3.4

Those numbers, summing to 10.2 events/1000, might seem low to you, given that your total ER visit and inpatient admission rate/1000 is roughly 200-250 in total. However, this is directionally accurate, even though the actual numbers are proprietary.

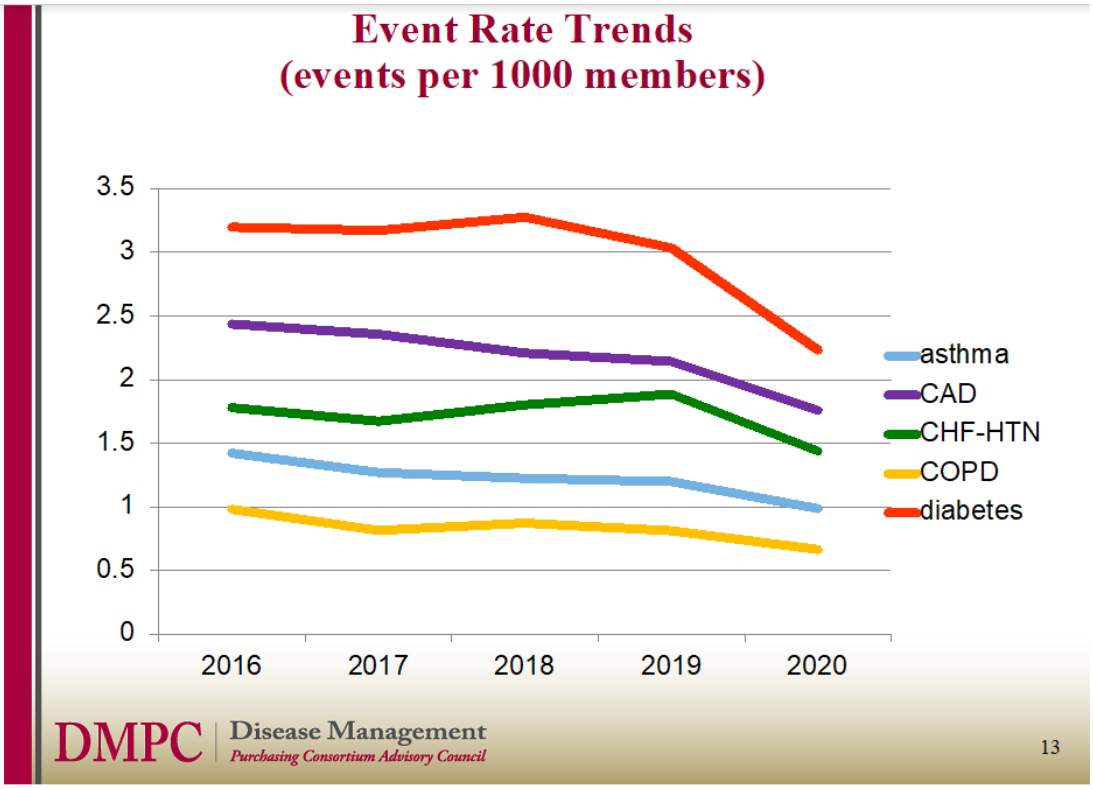
Adjusting for COVID

While a risk factor analysis would be valid year-over-year, most claims fell dramatically in 2020. The hospital business likely won't return to "normal" until 2022. That means you can't use 2019, 2020 or 2021 as a baseline for anything other than risk factors (and engagement – to follow below).

So trying to baseline a comprehensive disease management program would not be possible. However, if you want to do a program for diabetes, or a wellness program covering the three cardiometabolic conditions on the list, you could use the event rates for the other conditions as a baseline. Here is an article on a research project which did exactly that, well before the pandemic. Holding the admission and ER visit rates for the other common chronic conditions constant revealed a differential 6% reduction in diabetes adverse event rates as compared to the trends in the others.

Here is a sample tally of event rates from a large group showing that even for non-elective care – the wellness-sensitive medical events – utilization of emergency and inpatient declined quite a bit from the trend of the previous 4 years. The declines were roughly in the 25% range, after 4 years in which no event rate changed up or down more than 5% from the prior year.

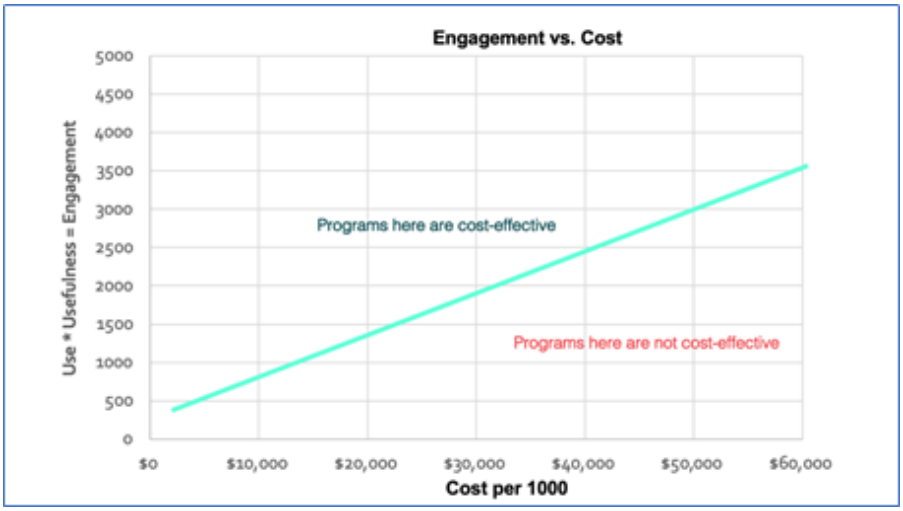
You could not attribute this gross decline to any program to keep people out of the hospital.



Doing an RFP for a program promising engagement

The elegance of the Validation Institute's Benefits Engagement Survey Tool, as described in Part Six, means that designing an RFP for such a tool should be easy, and it is.

Return to the baseline template from Part Six:



Link the RFP questions to the Benefits Survey Evaluation Tool. As with the clinical outcomes, be as specific as possible in naming:

(1) The other engagement tools you will be measuring them against
(2) The length of time following implementation when you will be measuring
(3) The approximate cost of the other tools

The expectation is to be above the line of best fit for the set of tools as a whole. As in this brief video, a vendor should be willing to put a substantial amount of fees at risk that it will be above the average of the other tools. Use a variation of the outcomes table to create a table measuring the cost-effectiveness of the engagement tool being evaluated.

Percent of fees at risk	Cost-effectiveness vs. average engagement tool		
	+20%	+40%	+60%
20%	Price	Price	Price
40%	Price	Price	Price
60%	Price	Price	Price
80%	Price	Price	Price

As you can tell from the sample goals in the table, it is much easier to move the needle for employee engagement than for risk reduction or reduction in admissions. The difference among vendors and programs is also much greater, as Part Six shows. Measurement is also easier, using the tool in Part Six.

Key take-aways

Compare these RFP queries and structures to anything you've seen from consulting firms. It will then become clear that, starting with the question at the top of this posting ("What is the penetration of this solution in your own insured population?"), the important thing is to ask the right questions, not the most questions. You can use this information to do exactly that, or we can send you to an advisor that can help.