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

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?

Earlier installments of this series have shown how regression to the mean, participation bias, comparisons to "trend," and investigator bias can dramatically overstate savings. Reliance on these fallacies obviously discourage many organizations from applying for validation. For those that do apply but rely to some degree on these fallacies nonetheless, usually out of necessity, the Validation institute ("VI") evaluators spend a lot of time teasing out the actual program impact from the biases. That is an inexact science at best, and is why virtually all evaluations come with "Limitations."

Limitations or not, validated outcomes are real enough that the Validation Institute places up to $25,000 into a "Credibility Guarantee" that we stand behind those validations.

Now, however, the Consolidated Appropriations Act has changed the game for obtaining and using one's own data. As a result, estimations could be much more accurate.

So much more accurate that the Validation Institute is offering the option to increase the Credibility Guarantees to as much as $1,000,000 if organizations can present proof statements using a full data set from an entire organization.

Note that as with any insurance, the higher the coverage the higher the fee. Nonetheless, this option is now available, whereas the previous cap was $25,000. There is an automatic $50,000 guarantee that comes with each of these validations, as a baseline.

There are five methodologies that the VI would consider valid enough to qualify this highest level of validation, known informally as VIPrime.

Note that we are not picking winners and losers in any given segment using these methodologies, but rather any organization that meets this standard qualifies. And further, with up to $1,000,000 at stake, the validation language will be carefully crafted. For example, we might say:

1. "The most valid study to date in this segment to date," or even
2. "By a considerable margin the most valid study in this segment to date that we are aware of," and then if another study of comparable validity comes along, both might say:
3. "This is one of the two most valid studies in this segment."

Examples of valid parallel study designs

A parallel study is one where would-be participants are randomly assigned to control or the study group. The randomized control trial (RCT) is one such methodology. In the case of drugs, the control group gets a placebo, so they don't know whether they are getting the drug or not.

This is called a "blinded" study. In many drug studies, even the investigators don't know. This is called "double-blinded."

Neither is possible in population health because you would know whether you are in a wellness/diabetes/prenatal program or not. So the standard for validity is lower. Nonetheless, outcomes should be clear and valid in any sizable parallel study.

Even unblinded RCTs are rarely undertaken in population health because (in addition to employers not hitherto having access to claims data) such studies need extra layers of review before proceeding, as ERISA plans are required to offer the same benefits to every employee absent Institutional Review Board approval. (One easy way around this is to offer the intervention to all comers, but promote actively to some worksites but not others.)

An example of an RCT would be Sera Prognostics. In high-risk pregnancy, an RCT is straightforward to conduct. The endpoints and outcomes are well-established. You can read their example here. This impact on NICU length of stay would qualify for VIPrime, should they be so inclined. The validation would not say that prospectively an organization would achieve results similar to this one but rather that this result proves that significant declines in NICU days are possible.

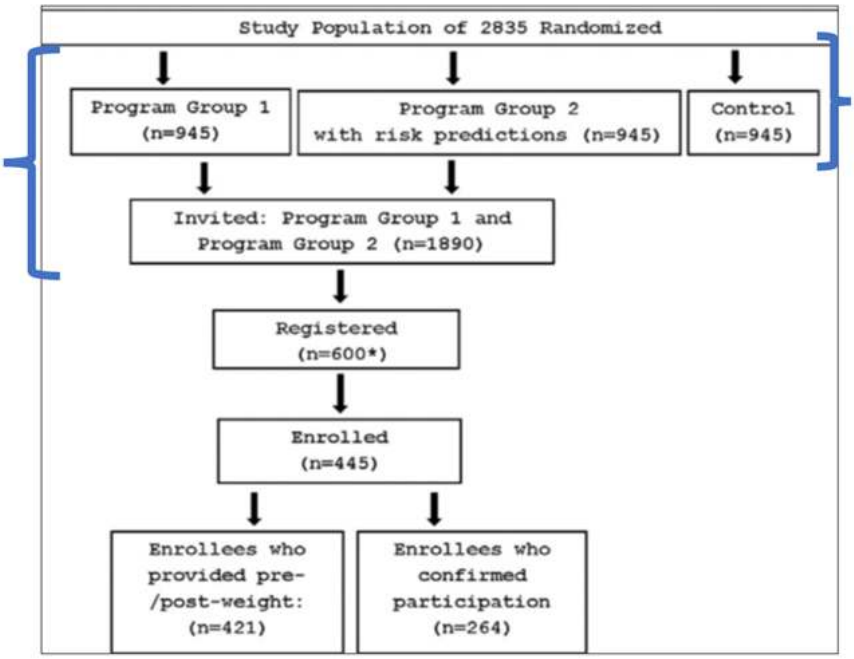
Often in employer populations, you need to separate people not just by willingness to participate, but also by site. Otherwise, there could be employee pushback. For example, in Virta's study, the parallel assignment took place in different sites, to minimize the possibility that (though the consent included the possibility of being in the control population) one diabetic employee might demand the intervention that the others are getting, once they see how helpful it is. Odds of this complication are minimized by putting the entire study population in one site while the control population is in the other.

In Sera's and Virta's cases, only the consenting pregnant women and diabetics were measured and compared. The consent took place before the assignment to control or study groups.

It would also be valid to have the consent take place after assignment to control or study groups

An example of parallel assignment that does exactly that would be cluster randomization, where entire sites are compared. Each site would have some employees who would consent to participate given the opportunity and some employees who wouldn't. The best example of a cluster randomization would be BJ's Wholesale Club. Wellness program participants were recruited at 20 sites, out of 160. Instead of comparing the recruits at those 20 sites to the entirety of the 140 control sites, or even to the non-participants at those 20 sites, the investigators compared the sum of those 20 sites in total to the other 140.

The best illustration of the importance of comparing like populations using parallel assignment is the following one. The population, having been qualified on the basis of claims data to have risk factors but not actual disease, was separated into three prima facie similar groups of 945 people before inviting the 1890 people in the first two groups to participate.



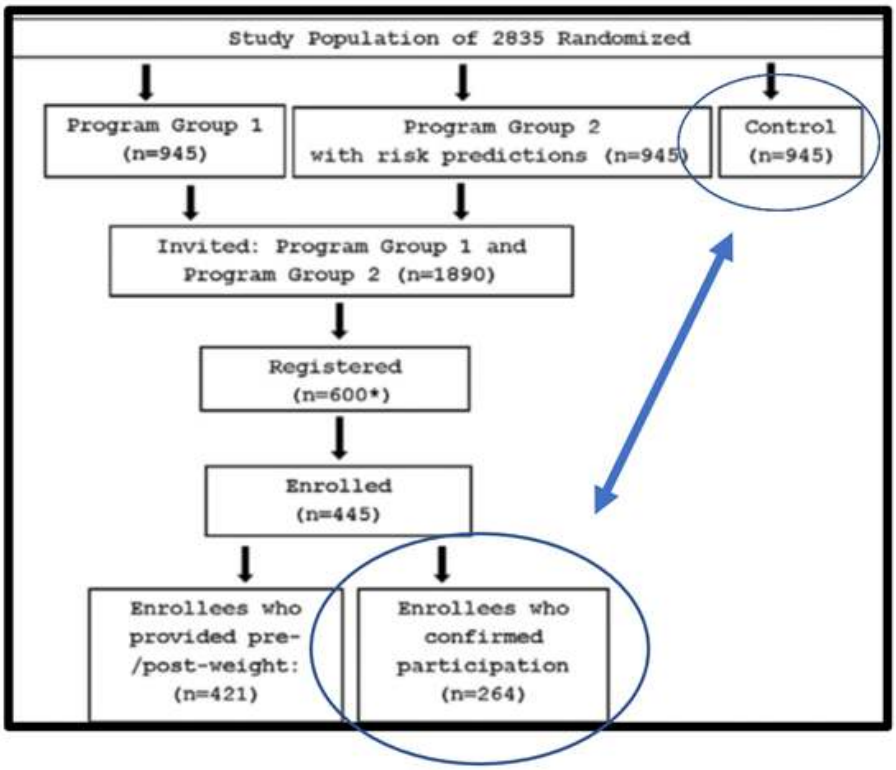
This is the RCT part

Following the intervention, investigators then compared the Invited Program Groups 1 and 2 to the Control Group, and found that basically nothing happened. Only one risk factor improved by a (barely) statistically significant amount. And three of the six health status indicators deteriorated vs. the control group.

Clinical Measure	Control ($N=945$)*	Invited (Groups 1 and 2 Combined) ($N=1890$)*	P
Waist (inches)	−0.48 ($n=653$)	−0.77 ($n=1311$)	0.06
Triglycerides (mg/dL)	−2.56 ($n=737$)	−8.12 ($n=1477$)	**0.05**
HDL (mg/dL)	1.44 ($n=722$)	1.63 ($n=1475$)	0.60
Glucose (mg/dL)	−0.10 ($n=723$)	2.11 ($n=1499$)	0.08
SBP systolic (mm Hg)	−1.48 ($n=650$)	−0.97 ($n=1306$)	0.47
SBP diastolic (mm Hg)	−1.47 ($n=650$)	−0.86 ($n=1306$)	0.22

A positive result using this design would have qualified for this level of validation had outcomes been significantly better in the invited Program Groups, since the RCT design was as good as it gets. (Likewise, it would have qualified had the thesis been that the intervention was very limited value, since this result is proof of that.)

However, precisely because the design was so good and the results so bad, the vendor elected to relegate this valid RCT analysis to a footnote. Instead, the headline was a comparison of the 945-person control group to the 264 participants (out of 1890) who actually completed the program:



Needless to say, this completely invalid result showed massive savings in the 1st year. No easy feat given that these 264 people had no diagnosed disease to begin with.

	Control ($N = 945$)	Participants ($N = 264$)	P	% Change
ledical cost PMPM	$434	$312	<0.02	−28%
ıpatient cost PMPM	$94	$45	<0.02	−53%
)utpatient cost PMPM	$271	$229	0.03	−15%
:mergency cost PMPM	$32	$18	<0.01	−44%
'harmacy cost PMPM	$109	$127	0.61	16%

And that example illustrates why parallel studies qualify for the highest level of validation we offer..

Other Valid Study Designs: Natural Experiments

Another valid design, also used by Virta, is a "natural experiment," in which the ordinary course of something else altogether outside the control of the investigators creates two equivalent groups. The most famous natural experiment in population health is the Oregon Medicaid study. Medicaid was expanded there to a higher income level, but slots were limited. Oregon Medicaid was expanded to encompass a higher income level, but slots were limited. Residents who met the new qualification standards also had to enter a lottery if they wanted coverage. Some won, some lost. But all met eligibility standards and all wanted to participate, thus creating the two roughly equivalent groups. (The two-year finding was that being covered by Medicaid as opposed to being uninsured didn't appreciably change physical health status, but did quite dramatically reduce both depression and financial strain.)

In Virta's case, the Veterans Health Administration (VHA) signed on as a client, but with a limited budget that could not accommodate all who qualified and wanted to participate. Therefore, those who were wait-listed became the natural control group. The VHA, whatever its other controversies, excels at data collection, and was able to compare the results of the actual participants to the would-be participants. (The size of the group and length of the study to date would not qualify it for this highest level. However, as a confirming analysis, it helps to qualify Virta.)

Other Valid Study Results II: Event Rate Plausibility Tests

Achieving an actual overall decline in inpatient, ER or other events was the original thesis for validation by the Validation Institute, as described in Chapter 2 of Why Nobody Believes the Numbers. An event rate plausibility test differs from other claims of declines in events and procedures in that all events are counted equally. This includes events incurred by people no one knew were sick. It is a simple count of ER and IP events with specific ICD codes. It can also be done for procedures, high-cost tests, etc. In all cases, a simple count. The COVID disruption would be a limitation but would not by itself disqualify.

An excellent example was the Iowa Medicaid diabetes program for rural counties. It reduced diabetes admissions by 6%. As you read the write-up, you'll notice the controls for this claim are event rates both for urban counties for diabetes, and for non-diabetes chronic disease event rates in the same rural counties. This claim did not get validation by VI for the simple reason that VI did not exist at the time.

Currently, the best example of an organization validated for event rate reduction is Blue Cross of South Carolina. The control for their program provided to their largest accounts is their book of business that has not signed up for disease management. Event rates have fallen somewhat more in the former than the latter.

Other Valid Study Results III: Actual Declines in Cost

Quantum Health's "special sauce," giving it an advantage over its competitors and carriers, is the immediacy with which they intervene with patients with cost-saving are management, following or even anticipating a qualifying event. By contrast, carriers can take weeks or months to reach out following such an event. As a result, over Quantum's entire book of business since inception, claims in the first full year are lower on average than claims in the year preceding adoption.

Qualifying for this highest level of validation by proving actual declines in cost is harder than it would appear. An applicant would need substantial data, rather than just a couple of examples. Further it could not be a decline "vs. trend." That decline may qualify for validation in general if the weight of the evidence we review confirms it, but the willingness of VI to expose itself to high-value arbitration claims requires actual declines, across a large number of customers.

Quizzify has validation for actual declines in cost of emergency room (ER) visits. The "ER Sticker Shock Prevent Consent" has been repeatedly proven to keep most ER bills under $1000. This track record would definitely qualify Quizzify, for ER visit cost reduction. Note that it would not by itself qualify Quizzify as a whole.

Several alternative PBMs that have validation today could also qualify as generating an actual decline in cost, largely because the three PBMs controlling 90% of the market have such high prices.

Valid Engagement Study and Results

A vendor may decide to achieve validation for engagement. Basically, every vendor claims high engagement, as described in Part Six of this series, and some are able to present a valid argument to the VI, but to qualify for the Gold Standard and the million-dollar Credibility Guarantee, only the Benefits Engagement Survey Tool (BEST) qualifies. To this date, only Quizzify has achieved validation using this tool at all, in one site. But to qualify for the $1,000,000 Credibility Guarantee, this tool must be deployed and results reported in multiple employer settings, with a minimum of 100 responses each, and engagement using this tool must be offered as a contractual guarantee as well.

Logistics of the Highest Level

Beyond the initial $50,000, each additional $100,000 in coverage carries an $8000 charge and a 20% copay. There is no additional fee if indeed we have already validated the program in question. Otherwise standard validation charges apply.

Endnotes:

Page 3:

- *https://validationinstitute.com/validated-provider/sera-prognostics/*
- *https://jamanetwork.com/journals/jama/fullarticle/2730614?guestAccessKey=f67976b4-63b8-4369-983f-196774f9404e&utm_source=For_The_Media&utm_medium=referral&utm_campaign=ftm_links&utm_content=tfl&utm_term=041619*

Page 4:

- *https://thehealthcareblog.com/blog/2015/12/16/genetic-testing-the-new-frontier-of-wellness-madness/https://theysaidwhat.net/2017/05/08/springbuk-wants-employees-to-go-to-the-bathroom/*

Page 6:

- *https://www.nejm.org/doi/full/10.1056/nejmsa1212321*
- *https://www.amazon.com/Why-Nobody-Believes-Numbers-Distinguishing/dp/1118313186/ref=sr_1_1?crid=22QMFT7A4PXGZ&dchild=1&keywords=why+nobody+believes+the+numbers&qid=1634497332&s=books&sr=1-1*

Page 7:

- *https://thehealthcareblog.com/blog/2013/07/30/stop-the-presses-a-disease-management-program-worked/*
- *https://www.quizzify.com/post/a-collection-of-er-sticker-shock-prevent-consents*

Page 8

- *https://validationinstitute.com/wp-content/uploads/2021/04/Part-6-How-to-tell-if-your-vendors-claims-are-valid.pdf*
- *https://www.validationinstituteexchange.com/best*