Annex 2: On the predictive value of Sustainability Framework indices

I stated in the main document that some of the limitations to our predictive ability are "*excessive expectations about the level of prediction from scores on the SF, even if extraneous / unpredictable events are dismissed.*" This annex will look into this.

In each of the six components of the Sustainability Framework, we collect indicators, transform them into points on a 0-100 scale, and then aggregate them to form indices (one for each component).

Let's for a minute forget how susceptible we remain to Black Swans in internal local system processes and in the environment, and just consider the variables we are able to measure in the Sustainability Framework.

This annex will consider two things:

- Non-linearity in how individual indicators and indices contribute to the prospect for long-term sustained gain;
- What are reasonable expectations from a complex model like the Sustainability Framework about probabilities of success.

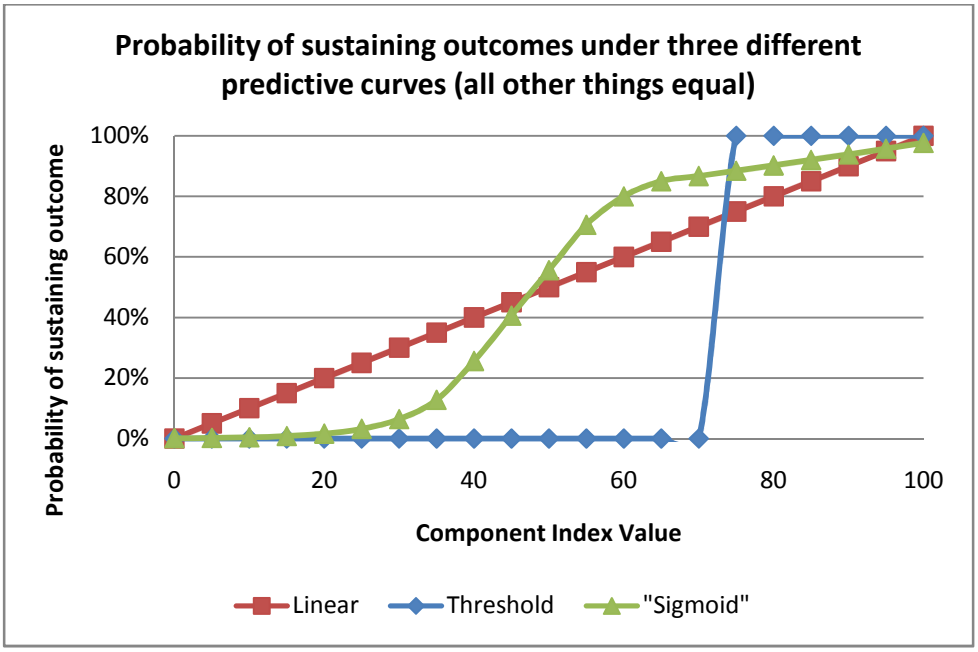
1- <u>Non-linear progression of indicators</u>

Let's consider a component index measure (although the same would apply to a single indicator) and assume for a moment that we're perfectly confident about all the indicators that go into it and its construction.

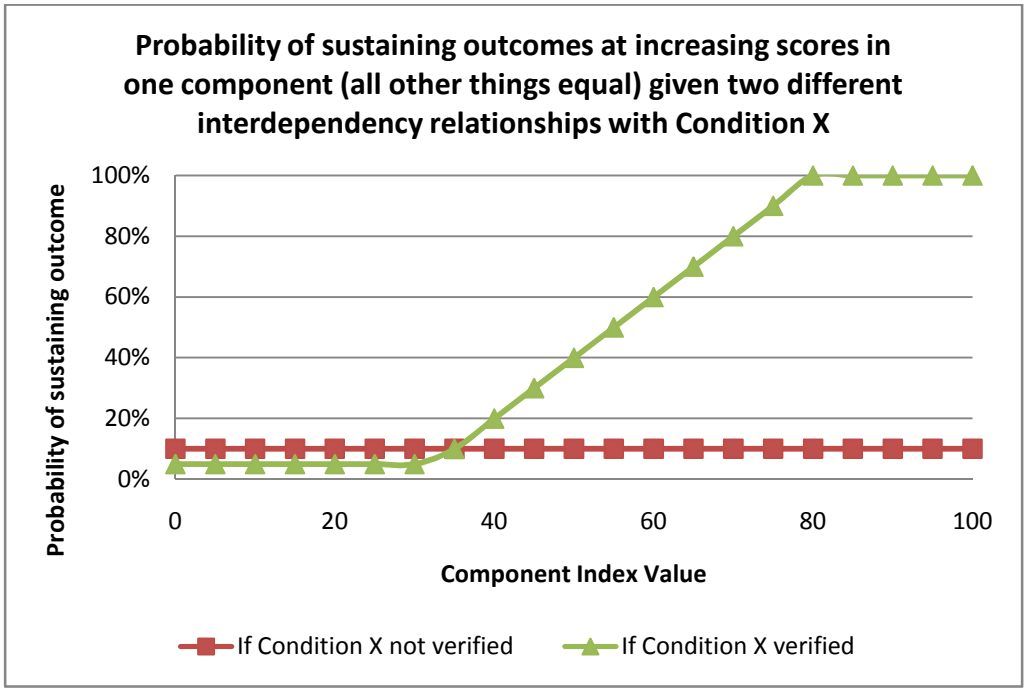
The nature of complex relationships between multiple variables means that—all other things being constant—assigning a predictive value to progress on this indicator can take different shapes. As an illustration the next figure shows three possible relationships between progress on this index value and the probability of sustaining outcomes over a set period of time. (Remember we hold all other things constant (*ceteris paribus*) and we have eliminated the unpredictable.)

The three patterns on the following figure are:

- Linear increase in probability of success (red line).
- "Threshold" progression, whereas probability remains nil until a threshold is reached, after which the probability quickly reaches certainty (100%).
- In between, the green line offers a progression close to a sigmoid curve.



The next figure simply shows an additional level of complexity. In this case we assume that if a certain condition is respected (Condition X) the predictive value of progress on a component measure is represented by the green line, while if Condition X is not respected, progress on the component being examined has no benefit and is represented by the red line.



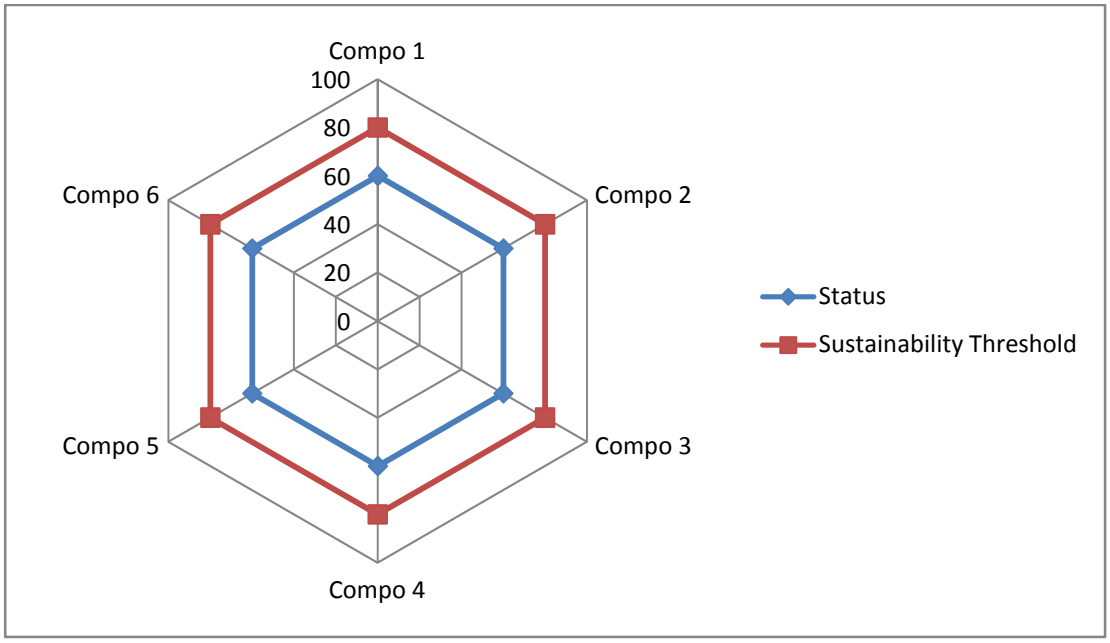
Such interdependencies are not hard to imagine. For example, what is the benefit of incremental gains in many areas of service quality, while commodities are still not available in the district facilities?

The point of all this is not to discourage the potential user of the Sustainability Framework, but to highlight the extent of the unknowns in our models and the necessity for trial-and-error to sort out where the thresholds are and what variables are really the drivers of any component. While we will have to remain humble about our ability to predict, we know that by not paying attention and not striving for progress on these different components, we have usually failed at the sustainability equation. Once again our failure to predict does not mean we cannot plan or prepare.

2- What are the implications of a multi-dimensional model on the probability of success

We just played with the concept of the predictive value of one measure with all other things being equal. Let's look now at the effect of having six components of evaluation.

The next figure presents a spider graph whereby each component index is above mid range – 60 – on a scale from 0 to 100. And let us consider that a score of 80 on all components is where we feel totally secure about the future prospects of the local system (results are strong in component 1 and above a theoretical 'sustainability threshold' on all others.)



Remember, we are still assuming a world without Black Swans; our model perfectly captures all relevant indicators, measures them perfectly and we have constructed perfectly valid index measures. We carried out our assessment and we're at the 60/100 in each component of assessment.[1]

It is very tempting to declare that we are, on average, at three quarters of the road to our sustainability threshold (60/80) and that consequently our chances for sustainability (i.e. maintaining or increasing the level of component 1 over a set period post-project) are 75%. But this would be a mistake for different reasons:

[1] Remember to treat this as a score (60 over a maximum score of 100), which is not the same thing as a percentage.

1- We do not know the nature of the relationship between our measures and the probability of success (see previous discussion), so a score of 60 could be much better or much worse than a 75% probability of success.
2- Even assuming the simplest relationship between a component index value and the probability of success (a linear relationship as in the first figure), we still have to deal with interdependencies between the components.

If a score of 60 for a single component represents a 75% chance of success—all other constraints withheld—then the probability of success is determined by probability of success based on Component 1 [p(C1)] AND p(C2) AND p(C3) AND p(C4) AND p(C5) AND p(C6).

The resulting probability is not 75% but (.75)* (.75)* (.75)* (.75)* (.75)* (.75) = $(.75)^6$ = 18%

This is a sobering number, but the corollary is that if all scores were half as much, the resulting probability would be 0.12%. This makes sense: what would you assess as the chances of seeing sustainably healthy kids five years from the end of a project which only scored 30 out of 100 on the health status component[2], the quality of services, the capacity of partners and communities and the enabling environment? Fairly low, I am sure. And improving our chances of sustaining positive results 150 fold (from a fraction of 1% to 18%) should be enticing to investors.[3]

Coming back to our main point—the issue is not that this 18% probability of success is correct and 75% is not; the point is that we should be cautious with the models we use if we want to speak of probabilities and prediction. This does not mean that we shy away from using models; rather, that we state our assumptions up front and reiterate them at the end, but in between we use our models to construct or derive or describe an evidence-based reality which then launches further learning.

Certainly the linear relationship between score and probability of success seems the most unlikely. We can pretty much assume that there are thresholds, that probability of success remains very low until these thresholds are reached, and that there is an acceleration in the increase of probability after these thresholds. Finally interdependencies and synergistic effects between elements of the model are certainly present. The predictive response on each axis may also vary.

Only empirically will we find what ultimately makes a difference and what we should focus on is planning and preparedness on related health objectives. Not being able to predict with precision should not keep us from improving the odds by making progress across all components.

[2] The health status which we want to sustain is itself a determinant of its own sustainability. Bossert was the first to show this in 1990, and this was found in other studies of determinants of sustainability. It makes sense really. Who would want to sustain 30% immunization coverage? Who would care to sustain coverage with ITNs before they start to show environmental benefits for all villagers?

[3] In an unpublished paper, we are showing that this kind of improvement in odds, applied strategically over multiple funding cycles can lead to 2 to 5 times the impact for the same dollar investment.