A method for estimating adult consumption effects of interventions for which we do not have direct evidence

Stephan J. Guyenet
GiveWell senior fellow

Summary

Development effects model spreadsheet
Copy of cost-effectiveness analysis with model input applied to VAS
Research supporting model scoring for SMC and VAS

The impact of childhood interventions on adult consumption ("development effects") is an important contributor to the cost-effectiveness of four of our current top charity interventions: insecticide-treated bed nets, seasonal malaria chemoprevention (SMC), deworming, and vitamin A supplementation (VAS). We have direct evidence on the impact of childhood SMC, insecticide-treated bed nets, and deworming on consumption in adulthood, but we do not have direct evidence on the adult consumption effects of VAS. Despite this lack of evidence for VAS, due to the findings on malaria and deworming, we believe that interventions that improve health during key developmental periods may generally have development effects. We currently model the development effects of VAS using an expedient method that is based on a simple analogy with anti-malaria interventions.

In this report, I describe a new method that uses indirect evidence to estimate the long-run effects of childhood VAS on adult consumption. A key assumption of this model is that certain short-term measures of the impact of an intervention in childhood are likely correlated with long-term effects on adult consumption. We chose measures that broadly reflect the quality of the developmental environment. The model is intended to be applicable to any intervention for which we do not have direct evidence on adult consumption effects.

In brief, the model works as follows:

- Collect evidence on seven variables that we believe may correlate with long-term consumption effects: direct evidence for adult consumption effects, cognitive ability, weight gain, all-cause mortality, morbidity, anemia, and plausibility. Weight these variables according to our beliefs about their relative importance.

- Semi-quantitatively score the effect size and certainty level of each variable for SMC and VAS using a structured scoring rubric.
- Calculate a weighted average of scores for SMC and VAS individually, and take the ratio of the two.
- Use this ratio to scale the absolute development benefit per person-year of SMC from our cost-effectiveness analysis, yielding an estimate of the development benefit of VAS or other interventions.

I applied this model to VAS. It estimates that VAS yields 47 percent as many units of value from development effects as SMC, per person-year of treatment.

The sections below contain details on the rationale for selecting and weighting variables, the rationale for scoring thresholds, and the application of this model to VAS.

Key remaining questions include:

- Should we include direct evidence as a variable in the model?
- Should we adjust for the external validity of evidence prior to scoring interventions?
- Does the model require too many variables (seven)? Should we pare it back?

Background and objectives

The impact of childhood interventions on adult consumption ("development effects") is an important contributor to the cost-effectiveness of four of our current top charity interventions: insecticide-treated bed nets, SMC, deworming, and VAS.[1] We model the development effects of the first three interventions using direct evidence linking the intervention to increased consumption in adulthood. Because direct evidence is not available, we currently model the development effects of VAS using an expedient method that is based on a simple analogy with anti-malaria interventions.[2]

Based on the findings of studies on malaria eradication[3] and deworming[4], and general plausibility,[5] we believe it is likely that interventions that substantially improve health in childhood tend to increase adult consumption. However, we do not have direct evidence on this for most of our interventions of interest, including VAS.

---

[1] See our cost effectiveness analyses of insecticide-treated bed nets, seasonal malaria chemoprevention, deworming, and vitamin A supplementation.
[2] This document explains our method.
[3] Bleakley et al. 2010 and Cutler et al. 2010.
[4] As-yet unpublished findings of the Kenyan Life Panel Survey, round 4, which examine the long-run impacts of a deworming RCT.
[5] It seems plausible that conditions that impose large health burdens during key periods of development adversely impact development, and that averting these conditions improve development, leading to improved adult outcomes.

We have two objectives for this work:

1. Refine our estimate of the development effects of VAS
2. Create a general development effects model that can be used to estimate the development effects of other interventions for which we do not have direct evidence

To estimate the development effects of interventions for which we have no direct evidence, we created a model that incorporates measures that we believe may be correlated with development effects. The model uses data on SMC and VAS to estimate the development benefits of VAS, but we intend it to be applicable to other interventions.

The development effects model

The model is based on the assumption that certain short-term measures of the impact of an intervention in childhood are likely correlated with long-term effects on adult consumption. We chose to develop a semi-quantitative model over a fully quantitative model because it is less time-intensive and allows us to incorporate qualitative information, such as biological plausibility and our uncertainty about quantitative evidence. The model works as follows:

- Collect evidence on the impact of the intervention on seven variables that we believe may correlate with long-term consumption effects. Weight these variables according to our beliefs about their relative importance.
    - 35 percent weight. Direct evidence for adult consumption effects
    - 15 percent weight. Cognitive ability
    - 15 percent weight. Weight gain
    - 10 percent weight. All-cause mortality
    - 10 percent weight. Morbidity
    - 7.5 percent weight. Anemia
    - 7.5 percent weight. Plausibility
- Semi-quantitatively score the effect size and certainty level of each variable for SMC and VAS using a structured scoring rubric. Scoring is on a 0-4 scale.
- Take a weighted average of scores for SMC and VAS individually, and take the ratio of the two.
- Use this number to scale the development benefit of SMC from our cost-effectiveness analysis (which is derived from direct evidence), yielding an estimate of the development benefit of VAS.
- This returns a figure for units of value per treatment-year from development benefits, which plugs in to the VAS cost-effectiveness analysis.

- All adjustments in the VAS cost-effectiveness analysis must then be applied to the figure: internal validity adjustment, external validity adjustment,[6] and coverage adjustment.

Rationale for the selection and weighting of variables

The model includes direct evidence for adult consumption effects, and indirect evidence from cognitive ability, weight gain, mortality, morbidity, anemia, and plausibility. These are variables that we believe are most likely to be correlated with effects on adult consumption. Most of them broadly reflect the quality of the developmental environment. Below, I discuss the rationale for each in greater detail.

We include direct evidence on the impact of the intervention on adult consumption because it is the single most informative type of evidence. Generally, if we are applying this method, we do not have direct evidence, or the direct evidence is very weak. Effectively, this variable penalizes the interventions we are assessing, pulling our overall estimate toward a skeptical prior. Since direct evidence is particularly informative, this variable receives the largest weight, at 35 percent. However, I am uncertain whether we should include this variable. If our prior is less skeptical, i.e. that interventions with similar short-term effects should produce similar long-term effects on consumption, including this variable and weighting it heavily will arbitrarily penalize interventions that have received less research on long-term outcomes.

We include evidence on the impact of the intervention on cognitive ability because cognitive ability appears to be one of the most important determinants of adult income. We estimate that in low-income settings, one additional IQ point yields 0.67 percent higher income.[7] For this reason, we weight it at 15 percent.

We include evidence on the impact of the intervention on weight gain for two reasons. First, growth is a broad measure of the quality of the developmental environment that integrates the impact of many stressors such as infection, undernutrition, and malnutrition. Second, physical size and robustness are probably important determinants of earning potential, particularly in low-income settings where the most common employment opportunities are manual. We chose weight gain rather than other measures of growth because it is commonly measured and may be a more sensitive measure of growth than height changes.[8] Due to its importance as a measure of the developmental environment and direct contributor to earning potential, we weight it at 15 percent.

---

[6] We apply internal and external validity adjustments after inputting the value into the cost-effectiveness analysis because most of the data used in the rubric come directly from unadjusted RCT results. Data on the effects of SMC and VAS are compared pre-adjustment in the rubric to make them more commensurate.

[7] This represents an 0.07 standard deviation increase in IQ. The rationale for the IQ-income relationship can be found in this report.

[8] See the “growth” section of this report for a discussion of SMC RCTs, which generally report that SMC increases weight but not height in children.

We include evidence on the impact of the intervention on all-cause mortality because it is a broad measure that integrates the impact of many stressors such as infection, undernutrition, and malnutrition. We include evidence on morbidity in years lost to disability (YLDs) for the same reason. We weight both at 10 percent due to their importance as measures of the developmental environment.

We include evidence on the impact of the intervention on anemia because anemia and/or iron status have a causal impact on cognitive ability.[9] Anemia is also commonly caused by chronic infections such as malaria, so it is a fairly broad reflection of infection status.[10] Since anemia is a narrower measure of health than growth, morbidity, and mortality, we weight it less at 7.5 percent.

We include plausibility to allow for information from the broader scientific literature, including animal research and biological mechanisms. For example, a severe form of malaria called cerebral malaria affects the brain in a minority of infected people, often causing coma, seizures, and lasting cognitive impairment.[11] This increases the plausibility that malaria impairs long-term cognitive ability in some people, and that malaria prevention measures increase average cognitive ability in a treated population. Since plausibility tends to be more subjective and uncertain than other measures, we weight it less at 7.5 percent.

Rationale for scoring thresholds

Scoring incorporates both the effect size and uncertainty of evidence for each variable. Effect size is quantitative, while uncertainty is a qualitative judgment. To score uncertainty, evidence is judged as “compelling” or “uncertain”.

- An example of “compelling” evidence is a well-conducted RCT, or meta-analysis of RCTs, providing direct evidence on the variable of interest.
- An example of “uncertain” evidence is one or more observational studies that provide direct evidence on the variable of interest and whose findings could plausibly be biased

[9] “Iron supplementation improved global cognitive scores (standardized mean difference 0.50, 95% confidence interval [CI] 0.11 to 0.90, p = 0.01), intelligence quotient among anemic children (mean difference 4.55, 95% CI 0.16 to 8.94, p = 0.04) and measures of attention and concentration.” Low et al. 2013, abstract.

[10] “Traditionally, the prevalence of anemia was used to estimate the prevalence of iron deficiency and IDA. However, in many developing countries, anemia can also result from infections such as malaria (5), from chronic inflammatory disorders (6), or from other nutritional deficiencies of folate or vitamin B-12 or A (7, 8). In addition, it is well known that infection and inflammation influence hemoglobin and iron-status indexes such as ZP and SF (9, 10) and, thereby, obscure the detection of iron deficiency.” Asobayire et al. 2001.

[11] “Cerebral malaria is the most severe neurological complication of infection with Plasmodium falciparum malaria. It is a clinical syndrome characterized by coma and asexual forms of the parasite on peripheral blood smears. Mortality is high and some surviving patients sustain brain injury which manifest as long-term neuro-cognitive impairments.” Idro et al. 2010, Pg. 1.

by confounding. Another example is a RCT that provides direct evidence on the variable of interest but has limitations that are likely to seriously undermine its informativeness.
- Natural experiments could be “compelling” or “uncertain”, depending on the likelihood that their design effectively captures causal effects.

Direct evidence is scored as follows:

Score	Explanation
0	No substantial evidence of effect, or compelling evidence suggesting a precisely measured null effect
1	Uncertain evidence suggesting a small effect size (<0.5% consumption increase)
2	Compelling evidence suggesting a small effect size (<0.5% consumption increase), or uncertain evidence suggesting a moderate effect size (0.5 - 1% consumption increase)
3	Compelling evidence suggesting a moderate effect size (0.5 - 1% consumption increase), or uncertain evidence suggesting a large effect size (>1% consumption increase)
4	Compelling evidence suggesting a large effect size (>1% consumption increase)

Effect sizes are calibrated on the impact of seasonal malaria chemoprevention (SMC) on children in areas with high malaria burden, which, prior to adjustments, we estimate increases mean adult consumption by 0.53 percent per person-year of treatment.[12] This seems like a fairly large effect, but I ranked it as “moderate effect size” to leave room for the possibility of larger effects.

We also estimate that prior to adjustment, deworming increases adult consumption by 11 percent.[13] This is far above the “large effect size” threshold, possibly arguing that the threshold should be higher. However, after applying our internal validity adjustment, the effect size is 1.4 percent,[14] which is not far above the threshold for a large effect. This raises the possibility that we should apply internal validity adjustments to estimates prior to scoring them in the rubric, although this would significantly increase complexity.

Cognitive effects are scored as follows:

Score	Explanation
0	No substantial evidence of effect, or compelling evidence suggesting a precisely measured null effect

[12] We estimate that averting counterfactual malaria for one year increases adult income by 2.3 percent, and that 23 percent of treated people in our beneficiary context gain this benefit per year. Multiplying these figures yields 0.53 percent.
[13] See this cell of our cost-effectiveness analysis.
[14] The internal validity adjustment is 13 percent. See this cell of our cost-effectiveness analysis. Multiplying these figures yields 1.4 percent.

1	Uncertain evidence suggesting a small effect size (<0.2 SD)
2	Compelling evidence suggesting a small effect size (<0.2 SD), or uncertain evidence suggesting a moderate effect size (0.2 - 0.4 SD)
3	Compelling evidence suggesting a moderate effect size (0.2 - 0.4 SD), or uncertain evidence suggesting a large effect size (>0.4 SD)
4	Compelling evidence suggesting a large effect size (>0.4 SD)

Effect sizes are calibrated on the impact of iron supplementation on cognitive ability in people with anemia, which we estimate increases general cognitive ability by 0.3 standard deviations in children and 0.35 standard deviations in adults.[15] This seems like a fairly large effect, but I ranked it as “moderate effect size” to leave room for the possibility of larger effects.

Growth effects are scored as follows:

Score	Explanation
0	No substantial evidence of effect, or compelling evidence suggesting a precisely measured null effect
1	Uncertain evidence suggesting a small effect size (<0.4 SD weight gain)
2	Compelling evidence suggesting a small effect size (<0.4 SD weight gain), or uncertain evidence suggesting a moderate effect size (0.4 - 0.8 SD weight gain)
3	Compelling evidence suggesting a moderate effect size (0.4 - 0.8 SD), or uncertain evidence suggesting a large effect size
4	Compelling evidence suggesting a large effect size (>0.8 SD weight gain)

Effect sizes are calibrated on the impact of ready-to-eat therapeutic foods on children with wasting (severe underweight), which we roughly estimate increase weight-for-height z-score by 1.1.[16] This seems close to the largest impact achievable due to the nature of the condition and intervention (giving nutritious food to young children who are underweight in part due to insufficient and/or low-quality food), so I ranked it as “large effect size”. Weight gain can be expressed as weight-for-age z-score change or weight-for-height z-score change.

All-cause mortality effects are scored as follows:

Score	Explanation
0	No substantial evidence of effect, or compelling evidence suggesting a precisely measured null effect
1	Uncertain evidence suggesting a small effect size (<10% reduction)

[15] See this report.
[16] See this row of our cost-effectiveness analysis of community-based management of acute malnutrition.

2	Compelling evidence suggesting a small effect size (<10% reduction), or uncertain evidence suggesting a moderate effect size (10-20% reduction)
3	Compelling evidence suggesting a moderate effect size (10 - 20% reduction), or uncertain evidence suggesting a large effect size (>20% reduction)
4	Compelling evidence suggesting a large effect size (>20% reduction)

Effect sizes are calibrated on the reduction of all-cause mortality caused by vitamin A supplementation in RCTs, which is 24 percent in a random-effects meta-analysis.[17] This seems close to the largest impact achievable due to the high baseline mortality burden and effectiveness of the intervention in these trials, so I ranked it as “large effect size”.

Morbidity effects are scored as follows:

Score	Explanation
0	No substantial evidence of effect, or compelling evidence suggesting a precisely measured null effect
1	Uncertain evidence suggesting a small effect size (<200 YLDs per 100,000 person-years)
2	Compelling evidence suggesting a small effect size (<200 YLDs per 100,000 person-years), or uncertain evidence suggesting a moderate effect size (200-400 YLDs per 100,000 person-years)
3	Compelling evidence suggesting a moderate effect size (200-400 YLDs per 100,000 person-years), or uncertain evidence suggesting a large effect size (>400 YLDs per 100,000 person-years)
4	Compelling evidence suggesting a large effect size (>400 YLDs per 100,000 person-years)

YLD values are drawn from Global Burden of Disease data. Effect sizes are calibrated on the reduction of morbidity caused by vitamin A supplementation and seasonal malaria chemoprevention (SMC), which we estimate avert 239 and 304 YLDs per 100,000 person-years of treatment, respectively.[18] This should be a fairly large morbidity effect because these interventions effectively target key causes of morbidity in beneficiaries. However, it seems conceivable that other interventions might have greater effects on morbidity, so I ranked them as “moderate effect size”.

Anemia effects are scored as follows:

[17] “At longest follow-up, there was a 12% observed reduction in the risk of all-cause mortality for vitamin A compared with control using a fixed-effect model (risk ratio (RR) 0.88, 95% confidence interval (CI) 0.83 to 0.93; high-quality evidence). This result was sensitive to choice of model, and a random-effects meta-analysis showed a different summary estimate (24% reduction: RR 0.76, 95% CI 0.66 to 0.88); however, the confidence intervals overlapped with that of the fixed-effect model.” Imdad et al. 2017, abstract.

[18] See “morbidity” section of this report.

Score	Explanation
0	No substantial evidence of effect, or compelling evidence suggesting a precisely measured null effect
1	Uncertain evidence suggesting a small effect size (<20% reduction in anemia prevalence)
2	Compelling evidence suggesting a small effect size (<20% reduction in anemia prevalence), or uncertain evidence suggesting a moderate effect size (20-40% reduction in anemia prevalence)
3	Compelling evidence suggesting a moderate effect size (20-40% reduction in anemia prevalence), or uncertain evidence suggesting a large effect size (>40% reduction in anemia prevalence)
4	Compelling evidence suggesting a large effect size (>40% reduction in anemia prevalence)

Effect sizes are calibrated on the reduction of anemia prevalence caused by iron supplementation in children. A meta-analysis of RCTs suggests that iron supplementation reduces the prevalence of anemia by 50 percent relative to controls.[19] Since iron deficiency is one of the primary causes of anemia and iron supplementation corrects it, this is probably the largest effect size achievable for an intervention, so I ranked it as “large effect size”.

Plausibility is scored as follows:

Score	Explanation
0	Very low plausibility; a meaningful effect seems unlikely
1	Low plausibility; a meaningful effect seems neither likely nor unlikely
2	Moderate plausibility; a meaningful effect seems somewhat likely
3	High plausibility; a meaningful effect seems likely
4	Very high plausibility; a meaningful effect seems very likely

Plausibility is the most qualitative metric and does not incorporate effect size thresholds.

Application of the model to vitamin A supplementation

A complete discussion of the evidence underlying my scoring of VAS and SMC can be found in this report. I applied the model in the following way to estimate the adult consumption effects of VAS:

[19] “Iron supplementation reduced the risk of anemia by 50% and the risk of iron deficiency by 79%.” Low et al. 2013, abstract.

1. Scored VAS and SMC across all seven variables in the model. Scoring can be found here.
    a. Direct evidence for adult consumption effects. I assigned SMC a score of 2 ("uncertain evidence suggesting a moderate effect size") because, prior to adjustments, we estimate that it increases adult consumption by 0.53 percent,[20] and the underlying evidence comes from natural experiments that have substantial limitations.[21] I assigned VAS a score of 0 ("no substantial evidence of effect") because I was unable to find direct evidence on its possible impact on adult consumption.[22]
    b. Cognitive ability. I assigned SMC a score of 1 ("uncertain evidence suggesting a small effect size") because the RCT evidence is mixed, with some trials reporting improvement in measures of cognitive ability but the most informative trial reporting no effect.[23] I assigned VAS a score of 0 ("compelling evidence suggesting a precisely measured null effect") because none of the seven trials I identified reported a significant improvement in measures of cognitive ability, and pooling these studies yielded a precisely estimated null result.[24]
    c. Weight gain. I assigned SMC a score of 2 ("compelling evidence suggesting a small effect size") because RCTs report fairly consistently that SMC causes weight gain of ~0.23 standard deviations.[25] I assigned VAS a score of 0 ("compelling evidence suggesting a precisely measured null effect") because a meta-analysis suggests that it does not significantly impact weight, with a point estimate near zero and moderately tight confidence intervals.[26]
    d. All-cause mortality. I assigned SMC a score of 4 ("compelling evidence suggesting a large effect size") because on the basis of RCT evidence, we estimate that SMC reduces all-cause mortality by 20.3 percent.[27] I assigned VAS a score of 4 ("compelling evidence suggesting a large effect size") because a meta-analysis of RCTs suggests that it reduces all-cause mortality by 24 percent.[28]
    e. Morbidity. I assigned SMC a score of 3 ("compelling evidence suggesting a moderate effect size") based on a straightforward calculation from the YLD

[20] We estimate that averting counterfactual malaria for one year increases adult income by 2.3 percent, and that 23 percent of treated people in our beneficiary context gain this benefit per year. Multiplying these figures yields 0.53 percent.
[21] The results of Bleakley 2010 and Cutler et al. 2010 underlie our estimate of the adult consumption effects of averting malaria in childhood. These studies primarily averted *P. vivax*, whereas the most common and impactful form of malaria in sub-Saharan Africa where our SMC beneficiaries reside is *P. falciparum*. See WHO, World Malaria Report 2008, figure 3.4, Pg. 11. For additional discussion of the limitations of these studies, see the cell note in this cell of our cost-effectiveness analysis.
[22] See our VAS development effects report, "Direct evidence on adult consumption effects of VAS".
[23] See our VAS development effects report, "SMC".
[24] See our VAS development effects report, "Vitamin A supplementation" and "Pooled findings: Vitamin A".
[25] See our VAS development effects report, "Growth".
[26] See our VAS development effects report, "Growth".
[27] See our VAS development effects report, "Mortality".
[28] See our VAS development effects report, "Mortality".

burden of malaria and the percent reduction in malaria incidence caused by SMC (estimated to avert 239 YLDs per year, per 100,000 treated).[29] I assigned VAS the same score for the same reason (estimated to avert 304 YLDs per year, per 100,000 treated).[30]

f. Anemia. I assigned SMC a score of 1 ("uncertain evidence suggesting a small effect size") because a meta-analysis of RCTs reports a nonsignificant trend toward an 18 percent reduction in anemia prevalence with SMC treatment (relative risk 0.82, 95% CI 0.65 to 1.04).[31] I assigned VAS a score of 2 ("uncertain evidence suggesting a moderate effect size") because a meta-analysis of RCTs and nonrandomized trials reports a significant 26 percent reduction in anemia prevalence with VAS treatment, but the meta-analysis has enough caveats that the findings appear uncertain (relative risk 0.74, 95% CI 0.66 to 0.82).[32]

g. Plausibility. I assigned SMC a score of 3 ("high plausibility; a meaningful effect seems likely") because malaria imposes large health burdens during key periods of development in SMC beneficiary populations, and a minority of people experience "cerebral malaria" that correlates with a higher risk of long-term cognitive deficits.[33] I assigned VAS a score of 2 ("moderate plausibility; a meaningful effect seems somewhat likely") because vitamin A deficiency imposes large health burdens (primarily diarrhea and measles) during key periods of development in VAS beneficiary populations.

2. Calculated the weighted average of scores for SMC and VAS. This calculation can be found here.
3. Took the ratio of the weighted average for SMC and VAS, which is 0.47. This implies that VAS yields 47 percent as much development benefit as SMC, per person-year of treatment. This calculation can be found here.
4. Calculated SMC units of value from development benefits, per person-year of treatment, based on values in our cost-effectiveness analysis, which yields 0.13 units of value. This calculation can be found here.
5. Scaled SMC units of value from development benefits using the ratio calculated above, which yields 0.061 units of value. This calculation can be found here.
6. Used this figure as an input into our VAS cost-effectiveness analysis. An example sheet can be found here (cells that were changed are highlighted in yellow).

Although the method behind the estimate is different than what we currently use, the cost-effectiveness output is scarcely changed.

[29] See our VAS development effects report, "Morbidity".
[30] See our VAS development effects report, "Morbidity".
[31] See our VAS development effects report, "Biomarkers".
[32] The two caveats I identified are the inclusion of nonrandomized trials and the inclusion of trials that did not isolate the effects of VAS. See our VAS development effects report, "Biomarkers".
[33] See our VAS development effects report, "SMC".

Alternative models

We considered two models for estimating development effects:

- Absolute. The input into the model is the absolute units of value per treatment-year of SMC, and the output is the absolute units of value per treatment-year of VAS. This is the model described in this report.
- Relative. The input into the model is the percent of total SMC benefits that come from development benefits, and the output of the model is the percent of total VAS benefits that come from development benefits.

Two of three GiveWell employees who have reviewed these models find the absolute model to be more intuitive, while the third finds the relative model more intuitive.

The relative model has the advantage of being simpler to integrate into the cost-effectiveness analysis because it does not require adjustments such as internal and external validity adjustments. Since it is scaled based on mortality benefits, it inherits the adjustments of mortality benefits. In contrast, in the absolute model, the input must be adjusted in the cost-effectiveness analysis. In the VAS cost-effectiveness analysis, adjustments are the internal validity adjustment, the external validity adjustment, and the coverage adjustment.

The primary advantage of the absolute model is that it is not distorted by scaling to mortality. This is not a problem for interventions that cause a similar mortality benefit to SMC, but it can cause substantial distortion if the intervention causes greater or lesser mortality benefit. For example, if an intervention averts 25 percent as much mortality as SMC per person-year but yields similar estimated development benefits to SMC, the relative model will scale the absolute development benefit to be 25 percent that of SMC, underestimating its absolute benefit by 75 percent. In contrast, the absolute model will correctly yield an absolute development benefit equal to SMC.

After considering these strengths and weaknesses, we tentatively favor the absolute model.

We also briefly considered a third method described by David Rhys Bernard here. It has the advantage of being fully quantitative rather than semi-quantitative, but it is more mathematically intensive and requires assumptions that we will not be able to satisfy in most cases.

Remaining questions

- Should we include direct evidence as a variable in the model? If we are using this method, we have little or no direct evidence on the adult consumption benefits of an intervention. Therefore, including this variable serves to penalize interventions for this

lack of evidence. This is logical if we have a skeptical prior about the ability of interventions to impact adult consumption (independent of the indirect evidence already considered in the model), but if our prior is less skeptical then this may be inappropriate. It also arbitrarily penalizes interventions that have received less research.

- Should we adjust for the external validity of evidence prior to scoring interventions? Presumably, the most accurate way to use the scoring rubric is to score based on our best guess of true effect sizes.[34] This argues in favor of performing external validity adjustments prior to scoring interventions. However, this significantly increases the complexity and time required to apply the model, so we have tentatively decided to omit it.
- Does the model require too many variables? For example, is the marginal information value of including morbidity and anemia worth the additional complexity and time requirement? There may be cases where all seven variables are not available, and this may force us to arbitrarily penalize interventions that have been less researched.
- Is the absolute or relative model preferable? We have tentatively settled on the absolute model, but it is not unanimously preferred.

[34] One extreme example is that long-term follow-ups of a deworming RCT report that deworming increases adult consumption by 11 percent. We apply an external validity adjustment of 13 percent to this estimate, implying that our best guess of the true treatment effect is actually 1.4 percent. This is a case in which the raw finding may be misleading. I am not aware of other cases that are this extreme, however.