CONSTELLATION
FUND

READING A METRIC
An Introduction

HEA111	HEA111 - Doula services leading to reduced maternal depression
Equation	(# women giving birth) x (Q1: % avoided maternal depression) x (QALY loss avoided) x ($ QALY)
Metric Description	This metric estimates the impact of doula care on the rate of maternal depression. Doulas are childbirth professionals who have training in pregnancy health education, childbirth preparation, labor support, lactation counseling, and newborn care. They serve as specialized home visitors, providing home-based education and support during the last half of pregnancy and typically for at least 6 weeks postpartum. Doulas accompany laboring women to the hospital to provide comfort and emotional support and to offer postpartum help around breastfeeding and bonding. One of the key aspects of the involvement of doulas is that they provide emotional and other support by maintaining a "constant presence" throughout labor, providing specific labor support techniques and strategies, encouraging laboring women and their families, and facilitating communication between mothers and medical caregivers and advocate on behalf of mothers within the healthcare system.
Metric Components	Number of participants: Number of women giving birth. Q1: Percentage of participants avoiding postpartum anxiety or depression: [SELECT] This is estimated using the formula: Q= ((ES*Base%)/(1-Base%+(Base% * ES))) - Base% In this formula, ES is the odds ratio of experiencing postpartum depression given the services of a doula, and Base is the base prevalence of postpartum depression in Minnesota. The odds [0.21] (American xiety reported by unted, as we are assuming a single year of depression prevented. $ value per QALY: [50000]
Metric Notes	No discounting required.
References	Falconi, A. M., Bromfield, S. G., uc Tang, T., Malloy, D., Blanco, D., Susan Disciglio, R., & Winnie Chi, R. (2022). Doula care across the maternity care continuum and impact on maternal health: Evaluation of doula programs across three states using propensity score matching. EClinicalMedicine, 50, 101531. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9257331/ MN Department of Health. Retreived from https://www.health.state.mn.us/docs/communities/titlev/postpartum.pdf

1		
2		
3		
4	$ value per QALY	50000
5		
6		
7		
8		
Adjustment for third-party outcome	Direct Outcome	100%
Need to discount to PV?		1.00
	Benefit	$36,828

PV Tool

Third-party outcomes Factor Tool

Strength of Evidence

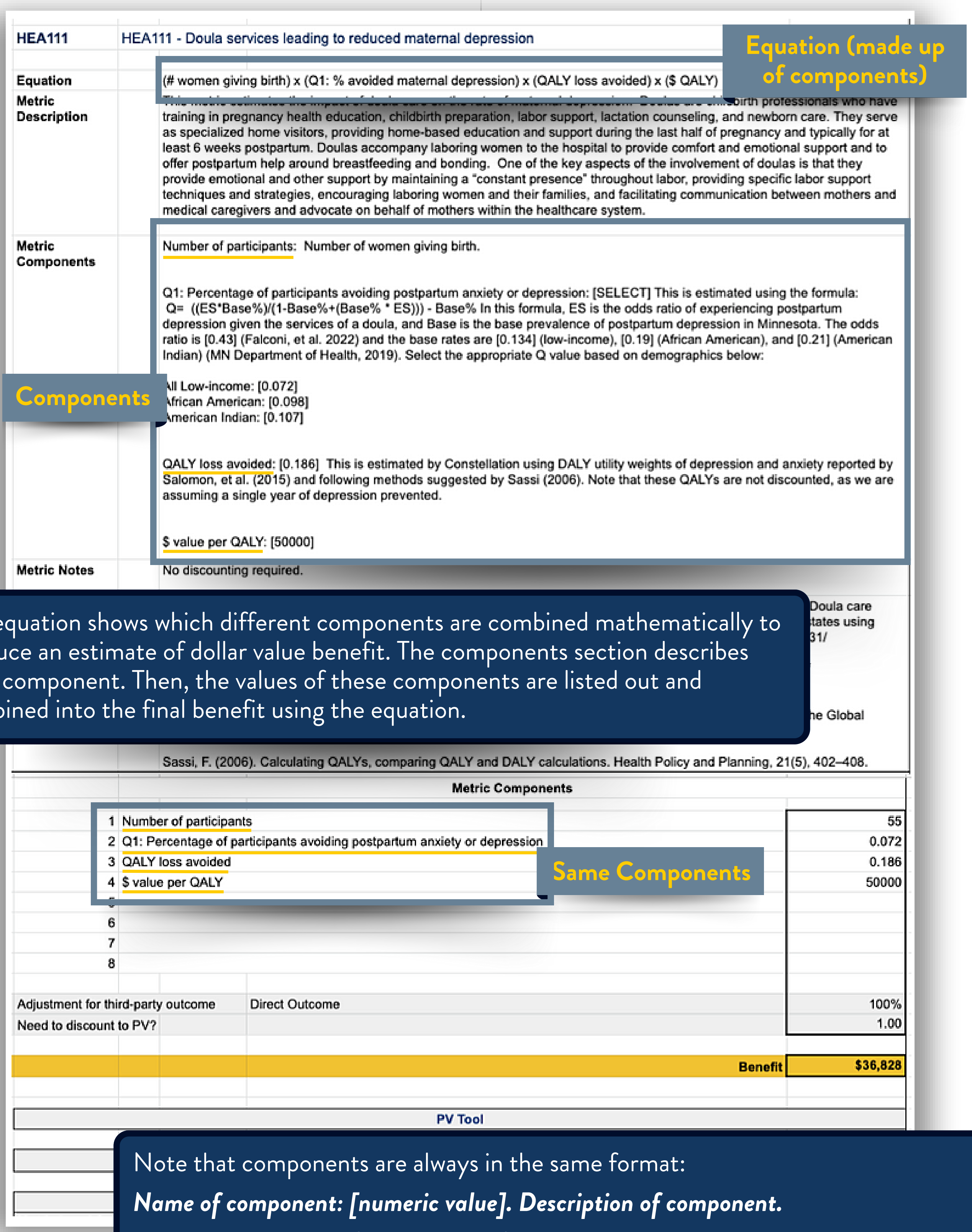
What is a metric?

When Constellation uses the word “metric”, we mean an equation that is used to produce an estimated monetary value of benefits for a certain activity.

What is this metric for?

The metric name gives you a basic idea of what activity is being evaluated (doula support in this case), and what outcome is being valued (maternal depression). The metric description describes the program as it is defined in the evidence.

How does this metric work?



HEA111	HEA111 - Doula services leading to reduced maternal depression
Equation	(# women giving birth) x (Q1: % avoided maternal depression) x (QALY loss avoided) x ($ QALY)
Metric Description	birth professionals who have training in pregnancy health education, childbirth preparation, labor support, lactation counseling, and newborn care. They serve as specialized home visitors, providing home-based education and support during the last half of pregnancy and typically for at least 6 weeks postpartum. Doulas accompany laboring women to the hospital to provide comfort and emotional support and to offer postpartum help around breastfeeding and bonding. One of the key aspects of the involvement of doulas is that they provide emotional and other support by maintaining a "constant presence" throughout labor, providing specific labor support techniques and strategies, encouraging laboring women and their families, and facilitating communication between mothers and medical caregivers and advocate on behalf of mothers within the healthcare system.
Metric Components	Number of participants: Number of women giving birth. Q1: Percentage of participants avoiding postpartum anxiety or depression: [SELECT] This is estimated using the formula: Q= ((ES*Base%)/(1-Base%+(Base% * ES))) - Base% In this formula, ES is the odds ratio of experiencing postpartum depression given the services of a doula, and Base is the base prevalence of postpartum depression in Minnesota. The odds ratio is [0.43] (Falconi, et al. 2022) and the base rates are [0.134] (low-income), [0.19] (African American), and [0.21] (American Indian) (MN Department of Health, 2019). Select the appropriate Q value based on demographics below: ll Low-income: [0.072] frican American: [0.098] merican Indian: [0.107] QALY loss avoided: [0.186] This is estimated by Constellation using DALY utility weights of depression and anxiety reported by Salomon, et al. (2015) and following methods suggested by Sassi (2006). Note that these QALYs are not discounted, as we are assuming a single year of depression prevented. $ value per QALY: [50000]
Metric Notes	No discounting required.

Doula care
tates using
31/

he Global

Sassi, F. (2006). Calculating QALYs, comparing QALY and DALY calculations. Health Policy and Planning, 21(5), 402–408.

Metric Components

1	Number of participants	55
2	Q1: Percentage of participants avoiding postpartum anxiety or depression	0.072
3	QALY loss avoided	0.186
4	$ value per QALY	50000
5		
6		
7		
8		

Adjustment for third-party outcome	Direct Outcome	100%
Need to discount to PV?		1.00
	Benefit	**$36,828**

PV Tool

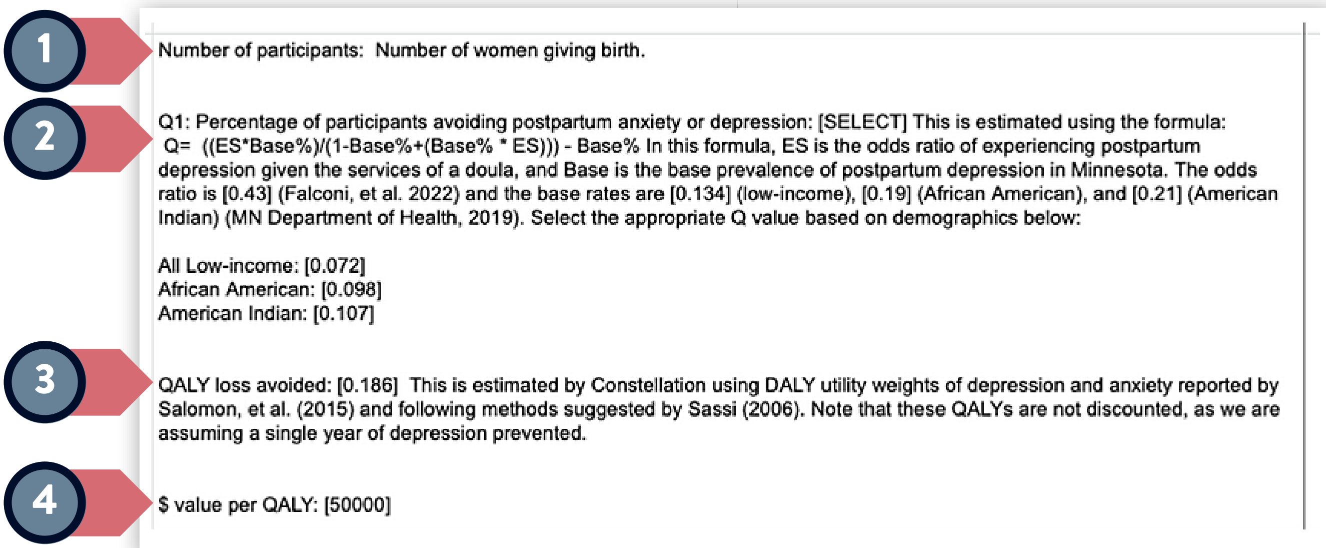
The equation shows which different components are combined mathematically to produce an estimate of dollar value benefit. The components section describes each component. Then, the values of these components are listed out and combined into the final benefit using the equation.

Note that components are always in the same format:

Name of component: [numeric value]. Description of component.

The name gives a brief description of the component. The numeric value is the actual number that will be used in the equation. The description goes into detail about where that number comes from.

What are components? What do they mean?



Number of participants: Number of women giving birth.

Q1: Percentage of participants avoiding postpartum anxiety or depression: [SELECT] This is estimated using the formula: $Q= ((ES*Base\%)/(1-Base\%+(Base\% * ES))) - Base\%$ In this formula, ES is the odds ratio of experiencing postpartum depression given the services of a doula, and Base is the base prevalence of postpartum depression in Minnesota. The odds ratio is [0.43] (Falconi, et al. 2022) and the base rates are [0.134] (low-income), [0.19] (African American), and [0.21] (American Indian) (MN Department of Health, 2019). Select the appropriate Q value based on demographics below:

All Low-income: [0.072]
African American: [0.098]
American Indian: [0.107]

QALY loss avoided: [0.186] This is estimated by Constellation using DALY utility weights of depression and anxiety reported by Salomon, et al. (2015) and following methods suggested by Sassi (2006). Note that these QALYs are not discounted, as we are assuming a single year of depression prevented.

$ value per QALY: [50000]

1

The first component in a metric is almost always the count of participants. This is usually data provided by you! This component may include a description specifying which participants should be used – in this case, the number of women giving birth.

2

Very often, metrics include one or more Q components – these are usually a percent value constructed from an effect size (more on this below). In this metric, this represents the percent of the total participants who avoid maternal depression as a result of the intervention. In this case, the [SELECT] is telling us that there are different options for this value. So, we might use 7.2%, 9.8%, or 10.7% depending on the demographics of participants.

3

Many of our metrics use QALYs to transform a health outcome into a dollar amount. QALY stands for Quality-Adjusted Life-Year and is a common way of evaluating a state of health in economics. One QALY is equivalent to one year lived in perfect health. 0.5 QALYs could mean half a year lived in perfect health, or one year lived at 50% of optimal health. In this case, avoiding postpartum depression means avoiding the loss of 0.186 QALYs.

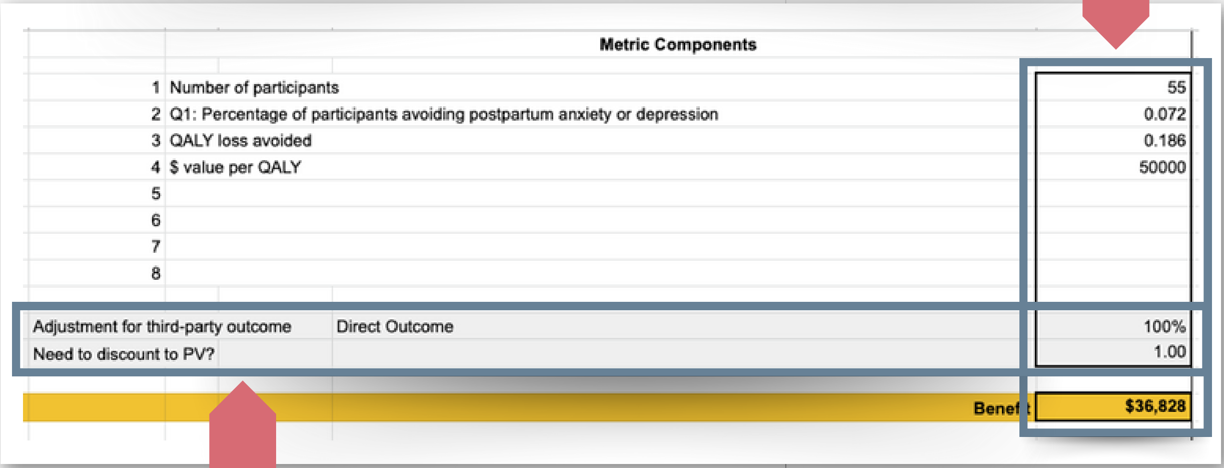
4

QALYs can be valued in a few different ways, but one common valuation – and the one that Constellation uses – is $50,000 per QALY.

Note: Benefits might be in the form of QALY or dollar gains OR in the form of avoided losses!

Putting it all together

This example metric estimates that providing doula services to 55 expectant mothers results in $36,828 worth of benefits by avoiding maternal depression. Note that this does NOT mean this is the ONLY benefit that this activity, support from doulas, creates. In the case of doulas, we have separate metrics to estimate benefits from reduced maternal mortality, and reduced low-birthweight. For any given intervention there may also be – and often are! – various positive outcomes that we are not able to build metrics around, either due to insufficient research available, or because the outcomes cannot be monetized.



	Metric Components		
1	Number of participants		55
2	Q1: Percentage of participants avoiding postpartum anxiety or depression		0.072
3	QALY loss avoided		0.186
4	$ value per QALY		50000
5			
6			
7			
8			
Adjustment for third-party outcome		Direct Outcome	100%
Need to discount to PV?			1.00
		Benef t	$36,828

What are these third-party outcome and PV discount adjustments?

PV discounting stands for present-value discounting - the economic discount to recognize that a dollar received today is valued more than a dollar received in the future. Some metrics project benefits that will accrue over a period of years. Often the benefit is already discounted in the metric as written, but when it isn't, Constellation discounts it here based on the time frame.

Next, many organizations will refer participants to other organizations for various services - for instance, imagine a clinic that provides primary care, but refers patients needing mental health support to a different organization. In these cases the second organization is providing the services that lead to benefits, but the first organization still deserves some credit for connecting the patients to that service. When this happens we apply the metric like usual, but include a referral adjustment. This can vary depending on how involved the first organization is in connecting the participant to the service, and on how certain the first organization is of how many referees indeed received services.

Into the weeds...

Wait, what is this Q-value thing? How do we know how many women avoid maternal depression?

Note: this section gets into the statistical concepts use in turning research results into estimates of benefits. If that doesn't sound fun to you, feel free to skip the next two pages!

Research papers may report their results in a large variety of ways. To write a metric, we often want to know the percent of a population who will have a certain outcome, or how impactful an intervention is on a given outcome. We have various mathematical tools available to help us 'translate' from whatever form a paper uses into a percent of total participants, or other value we can use in the metric. Let's explore a few of the most common ones.

Odds Ratios:

One common way of reporting dichotomous outcomes – e.g., things like a mother either does or does not develop postpartum depression – is with what's called an odds ratio. This is the ratio of the odds that someone who does receive the intervention develops postpartum depression, compared to the odds that someone who does not receive the intervention develops postpartum depression:

$$OR = \frac{\frac{P_1}{(1-P_1)}}{\frac{P_2}{(1-P_2)}} = \frac{P_1(1-P_2)}{P_2(1-P_1)}$$

Where P_1 is the percent of people who do receive the intervention but nonetheless develop postpartum depression, and P_2 is the percent of people who do not receive the intervention, and do develop postpartum depression.

In our example, notice that the odds ratio is less than one (0.43) – this tells us that someone who does receive the intervention (the services of a Doula) is less likely to develop postpartum depression.

But before we can say how many participants avoided postpartum depression, we need one more thing: to know how many would have developed postpartum depression without the program. To approximate this value, we use the prevalence (also called a baserate) of postpartum depression in Minnesota. This data is available by race and income, which is why we're able to calculate three different Q values for different populations.

Still in the weeds...

With this information, we calculate Q using the following formula:

$$Q_d = \frac{OR \times Base_{\%}}{1 - Base_{\%} + OR \times Base_{\%}} - Base_{\%}$$

Where OR is the odds ratio, and $\text{Base}_{\%}$ is the prevalence.

Standardized effect size:

For continuous outcomes – for instance, test scores – it's often helpful to use a standardized effect size. Researchers report standardized effect sizes so readers know generally how impactful or not a result is without having to be intimately familiar with the different measures involved. Suppose a tutoring program is shown to increase scores on a particular test by five points. This isn't really meaningful until we know "how good" of an improvement five points is. To think about this, we look at the standard deviation (SD) of test scores.

Standard deviation is a measure of how clustered or spread out data are. If the SD of test scores is 15, say, that tells us the range of scores is wide, and a 5 point increase isn't so impressive. On the other hand if the SD is 2, suddenly that 5 point increase is huge!

A standardized effect size is the difference a program makes divided by the standard deviation. The standardized effect size alone gives us a sense of how impactful a program is, but to use in our metrics, we often need to "translate" it back into its original units. To do this we multiply by the standard deviation of that measure within our population of interest - for instance, the SD of test scores among Twin Cities high school students. Standardized effect sizes are especially useful in combining results from various papers into a single meta-analysis. Suppose different studies report the effect of the same tutoring program on different test scores (with different standard deviations) – expressing the results as standardized effect sizes lets us combine them all into a single effect size we can then apply to the relevant metric.

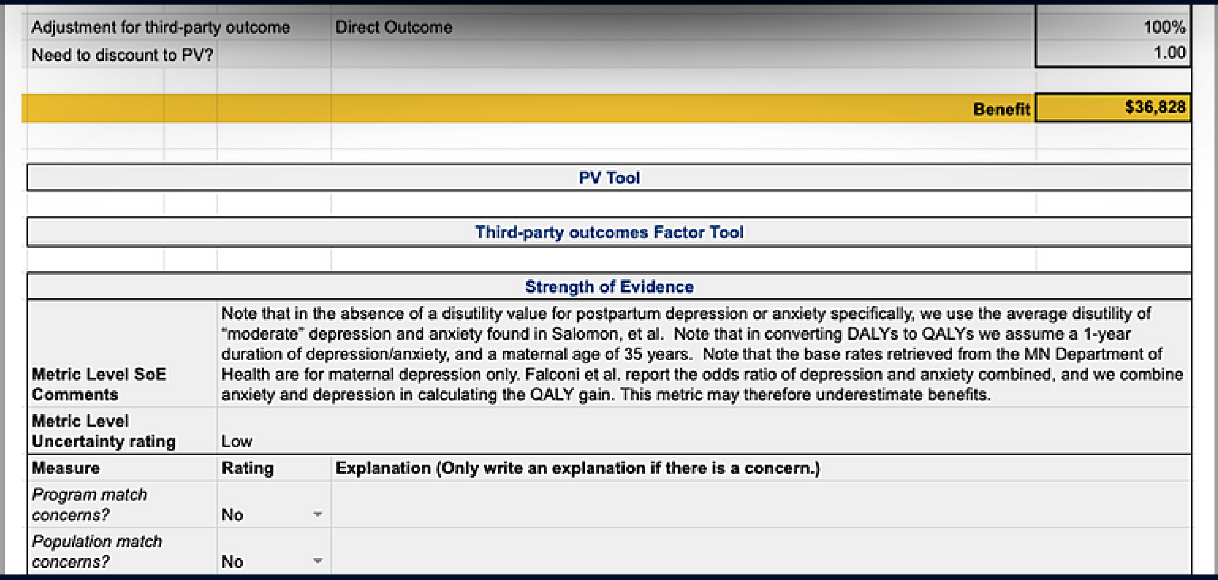
Phew! We're out of the weeds. Catch your breath and we'll get back to our regularly-scheduled programming.

Assumptions and strength-of-evidence

Metrics, and the calculations they facilitate, are meant to be reasonable estimates. Often in constructing a metric there might be multiple valid paths to take, or assumptions that could be made. Constellation chooses one approach, but it isn't necessarily the only valid approach. When interpreting the result of a metric, it's helpful to pay attention to what assumptions went into the metric, and where the evidence behind a metric might not be as strong as we'd like. These factors may be documented in the metric text itself, or in what we call Strength of Evidence notes.

Strength of Evidence notes contain anything from noting that a key study has a small sample size, or a sample that is likely to be demographically different from program participants, to documenting that we used a result that isn't quite the same as what we need but is a reasonable proxy. For instance, there are a few metrics where a program has an impact on out-of-home placement of children, and we use this as a proxy for child abuse. The two do not line up perfectly, but we have reason to believe they are related.

In this example metric, we make an assumption about the average severity of depression experienced by mothers experiencing maternal depression.



Adjustment for third-party outcome	Direct Outcome	100%
Need to discount to PV?		1.00
	Benefit	**$36,828**

PV Tool

Third-party outcomes Factor Tool

Strength of Evidence

Metric Level SoE Comments	Note that in the absence of a disutility value for postpartum depression or anxiety specifically, we use the average disutility of "moderate" depression and anxiety found in Salomon, et al. Note that in converting DALYs to QALYs we assume a 1-year duration of depression/anxiety, and a maternal age of 35 years. Note that the base rates retrieved from the MN Department of Health are for maternal depression only. Falconi et al. report the odds ratio of depression and anxiety combined, and we combine anxiety and depression in calculating the QALY gain. This metric may therefore underestimate benefits.	
Metric Level Uncertainty rating	Low	
Measure	**Rating**	**Explanation (Only write an explanation if there is a concern.)**
Program match concerns?	No	
Population match concerns?	No	

Constellation may also write Strength of Evidence notes related to the application of a particular metric for a particular evaluee. This is done when the data provided by the evaluee does not perfectly align with the data needs of a metric, or when the program offered by an evaluee is not exactly the same as the program estimated in the metric.

By its nature, we can't know exactly how much the issue noted in a Strength of Evidence note will impact the results a metric produces – if we did, we'd simply write it into the metric! – but by noting these important factors, we have a starting point for understanding the context of our estimate.

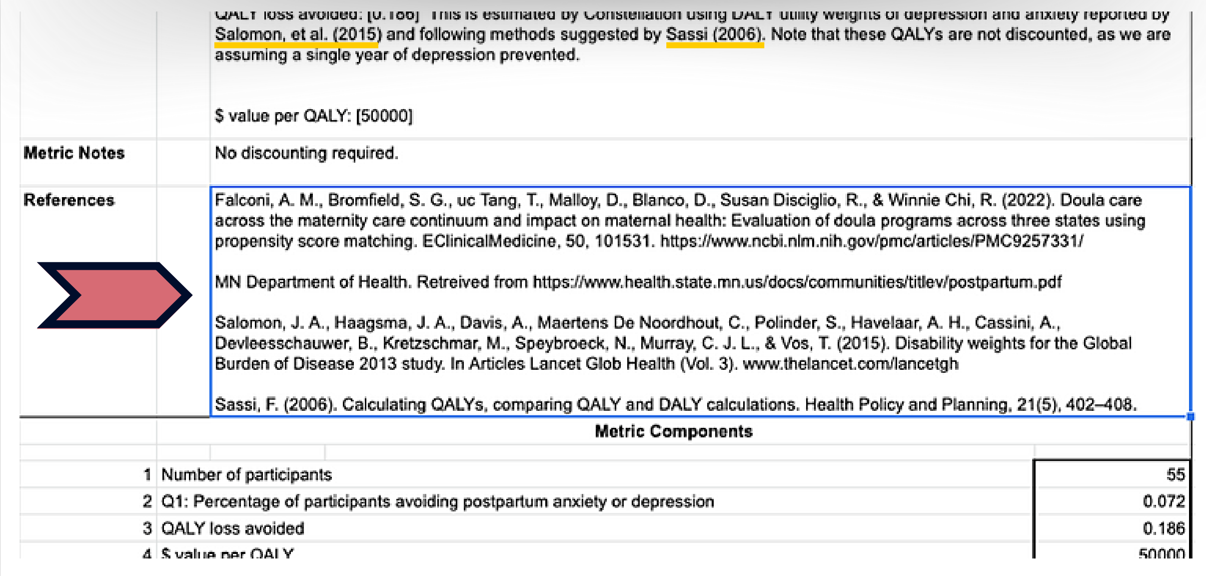
Final thoughts

What else should I look out for?

Remember, some programs may have various positive outcomes that Constellation does not have metrics for, either because of insufficient evidence, or because outcomes cannot be monetized. When considering the overall report, it's good to keep an eye out for which benefits of your work are not included. This is part of understanding what the BCR can - and cannot - tell you.

It can also be helpful to consider the overall "logic flow" of a metric. Sometimes the logic is very direct and straightforward – a health intervention, for instance, that improves patient health and thus delivers QALYs. But sometimes the route may be a bit more circuitous, especially if a more "direct" line to a benefit hasn't been sufficiently studied. For instance, a parenting program might be expected to improve a child's academic performance and thus lifetime earnings – but we may have to write a metric that connects the parenting program to reduced "problem" behavior in a child, and then connect that behavior to the likelihood of graduating from high school.

Note that whenever we use a piece of evidence in a metric – typically pulling a number from it – we cite the source briefly in the metric text and include the full citation in the citations section. This is a great place to start if you're curious about a piece of evidence or want to assess for yourself how good a fit it is to your program! Remember, a Constellation metric is only able to include monetizable outcomes – it could be the case that a piece of evidence also includes important information about a non-monetizable outcome!



	QALY loss avoided: [0.186] This is estimated by Constellation using DALY utility weights of depression and anxiety reported by Salomon, et al. (2015) and following methods suggested by Sassi (2006). Note that these QALYs are not discounted, as we are assuming a single year of depression prevented. $ value per QALY: [50000]
Metric Notes	No discounting required.
References	Falconi, A. M., Bromfield, S. G., uc Tang, T., Malloy, D., Blanco, D., Susan Disciglio, R., & Winnie Chi, R. (2022). Doula care across the maternity care continuum and impact on maternal health: Evaluation of doula programs across three states using propensity score matching. EClinicalMedicine, 50, 101531. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9257331/ MN Department of Health. Retreived from https://www.health.state.mn.us/docs/communities/titlev/postpartum.pdf Salomon, J. A., Haagsma, J. A., Davis, A., Maertens De Noordhout, C., Polinder, S., Havelaar, A. H., Cassini, A., Devleesschauwer, B., Kretzschmar, M., Speybroeck, N., Murray, C. J. L., & Vos, T. (2015). Disability weights for the Global Burden of Disease 2013 study. In Articles Lancet Glob Health (Vol. 3). www.thelancet.com/lancetgh Sassi, F. (2006). Calculating QALYs, comparing QALY and DALY calculations. Health Policy and Planning, 21(5), 402–408.

Metric Components

1	Number of participants	55
2	Q1: Percentage of participants avoiding postpartum anxiety or depression	0.072
3	QALY loss avoided	0.186
4	$ value per QALY	50000