EVIDENCE EVALUATION:

A guide for how to apply the Bridge Collaborative Evidence Rubric

Mac users see footnote*

Why it was created – and why, when, and how to use it

The evidence rubric presented here was created by the Bridge Collaborative to help assess and communicate the strength of confidence in evidence for hypotheses about the effects of interventions. In this context, the hypothesis is one that describes the relationship between two elements of a system (i.e. two nodes in a results chain). While the rubric and tips for use provided here could be applied in other contexts, the application to hypotheses about intervention effects was the context in which this was created and written.

Once available evidence has been collected, stating the confidence in the hypotheses reflected by each link can be useful in several ways[1]:

- First, reviewing evidence can improve the chain, often further resolving linkages and clarifying hypotheses and assumptions
- Looking at the confidence in a chain can inform risk assessment (where are areas of uncertainty and high risk?) and decisions about whether and how to proceed with an intervention.
- Use of the rubric can help make an argument about the relationship between two aspects of a system credible to external audiences.
- We introduce a shared evidence evaluation rubric because each sector currently has multiple rubrics for evaluating evidence. For consistent interpretation of evidence from multiple sectors, a common rubric is needed. The Bridge Collaborative has developed a rubric with elements of confidence that are shared across communities.

We do not prescribe a minimum amount of evidence that must be reviewed to be able to meaningfully use this rubric, however the thoroughness of the review should be aligned with the risk associated with the outcomes or action that the confidence determination will inform.

* This interactive PDF may not display or function properly in Mac Preview. Mac users are asked to view this worksheet in a PDF reader such as Adobe Acrobat Reader DC (free to download).

[1] This list is a modified excerpt from the Bridge Collaborative Practitioner's Guide

Rubric elements

Criteria:

The rubric is comprised of four columns that are used to determine the overall confidence level[2]:

- **Types of Evidence:** We take a broad and inclusive definition of evidence and accept information from the following sources: expert knowledge, measurement results, models, qualitative and quantitative studies, and theories.
- **Applicability:** The similarity in ecological, social, political, economic or other relevant conditions between those represented in the available evidence and those in the case to which the evidence is being applied.
- **Consistency:** The agreement across findings in a body of evidence, not the lack of variability in observed relationships.
- **Accepted methods:** Methods that have been peer reviewed and broadly supported by a community of practice.

Level of confidence:

The rubric includes four confidence levels: high, moderate, fair, and low[3].

While the language differentiating the levels may seem vague (e.g. "several" types of evidence versus "a few" types of evidence), initial testing of the rubric has indicated that these designations were useful once a specific set of evidence was in hand for evaluation.

In this rubric, high confidence in the hypothesis underlying a link can be stated when multiple types of evidence support the assumption; results are consistent across sources, types of evidence, and contexts; methods used across evidence types are well documented and accepted by the relevant field(s); and available evidence is highly applicable to the study or practice context. All four criteria must be met for a high confidence statement to be applied.

Other than for the 'high' category, we do not provide a hard rule for how to determine the overall confidence level based on the assessment of each column, but rather provide options and guidance (in 'Rubric Tips' below) for how your team may choose to do so.

[2] This list is a modified excerpt from the Bridge Collaborative Practitioner's Guide

[3] We introduce four levels rather than three for two reasons: 1) research shows a human cognitive bias towards a middle category when 3 options are provided, and the presence of a 4th category can help mitigate this effect, and 2) the 4 levels introduce additional information and nuance that would be lost with three categories.

Evidence Rubric

Confidence level	Criteria			
	Types of evidence	**Consistency of results**	**Methods**	**Applicability**
High	Multiple	AND consistent across sources, types of evidence, and contexts	AND well documented and accepted	AND high
Moderate	Several	Some consistency	Not fully accepted, some documentation	Some
Fair	A few	Limited consistency	Emerging, limited documentation	Limited
Low	Limited, extrapolations	Inconsistent	Poor documentation or untested	Limited to none

Unified evidence evaluation rubric for identifying confidence in results chain assumptions across health, development and environmental evidence. Applicability refers to the similarity in ecological, social, political, economic or other relevant conditions between those represented in the available evidence and those in the case to which the evidence is being applied. Consistency refers to the agreement across findings in a body of evidence, not the lack of variability in observed relationships. We define accepted methods as those that have been peer reviewed and broadly supported by a community of practice.

Rubric tips

The rubric was first released in 2017, and has since been applied by multiple teams working with the Bridge Collaborative. Below, we provide tips for using the rubric, based on their experience and feedback.

Getting started with the rubric:

Tip 1: **Start with a good hypothesis**: Each link in your results chain should represent a single, testable hypothesis. It is best to use the rubric to test a specific question. If you start with a focused and well-structured scientific question within a specified geographic context, using the rubric will be easier. Unspecified, generalized hypotheses often result in lower evidence confidence levels. If you need to answer a more generalized question, it will be more difficult to define where you fall on the applicability criteria, and results are more likely to be inconsistent (and context dependent).

Tip 2: **Terminology is key:** Sometimes the way you phrase a question will determine the evidence you find. Sometimes you need to alter your terminology to find the evidence you need. For example, a group was trying to examine the evidence for the results chain made up of the following links: sediment capture by marsh plants → suspended sediments in stream water → quality of drinking water. The group was having trouble finding evidence for the first link. It was only after searching for evidence using the term "accretion," rather than "sediment capture" that they were able to find relevant and useful evidence. Sometimes you need to reach out to an expert, even if it is within your own sector, to determine what key words might be to find the right evidence.

Tip 3: **Ask around:** Get expert advice to point you to relevant, credible evidence. This is a great way to be guided to relevant evidence and will help to shorten the amount of time required to find the evidence you need.

Assigning confidence levels:

Tip 1: **Define what each confidence level means for you**: Language in the rubric is purposefully left general so that each user can decide how to define what each confidence level means in specific terms for a particular application. For example, for the consistency criteria, each team using the rubric will need to define how they distinguish between consistent, some consistency, limited consistency, and inconsistency of results. This calibration provides an important opportunity to surface often hidden assumptions or differences in interpretation so that everyone participating in the assessment is on the same page.

A team may want to add requirements to each criteria to further define each confidence level. This is especially helpful if multiple people are working to grade different bodies of evidence that will be compared.

Tip 2: **Define how you will determine confidence levels consistently:** As stated above, we do not provide hard rules for how to determine an overall confidence level based on the assessment of each column. However, it is helpful to make a conscious choice about how you will be assigning confidence levels, given how the evidence scores in each column. A few potential approaches are described here:

Lowest common denominator: the overall confidence score is assigned as the level that corresponds with the lowest scoring column in the rubric. For example, if three columns meet the criteria for 'moderate', and one column meets the criteria for 'low', the overall confidence score is 'low'. (Note: In our experience to date, this is the option that teams have most often deemed appropriate for their planning context.)

Average: the overall confidence score is assigned based on a perceived average of where individual columns score. For example, if the evidence meets the criteria for two columns as 'high', and two columns as 'fair,' the overall score is 'moderate.'

Assigning numerical scores: You may choose to assign numerical scores for each confidence level, and then create a combined score for each body of evidence. The team then can choose cutoff scores that are assigned for each overall confidence level. For example, if we assign column scores where 'high'=4; 'moderate'=3; 'fair'=2; and 'low'=1, then an overall confidence score would be assigned where 16 = high, 9-15 = moderate, 6-9 = fair, and 0-5 = low. A body of evidence meeting the four component criteria at 'high,' 'high,' 'fair,' and 'fair' would have an overall score of 12 (4+4+2+2), giving it an overall confidence level of 'moderate.'

Tip 3: **Decide how you distinguish between 'moderate' and 'fair' confidence:** The two middle confidence levels are quite similar— but each user should be consistent in how they decide to distinguish the two. In our experience, links with moderate ratings are ones that would be high if not for one criterion. Having two weaker criteria or one very weak criterion is more likely to end up as fair. We have also found that for those who think consistency of evidence is crucial, if results lack consistency, they choose not to rank a link as moderate even if all of the other criteria are strong.

Tip 4: **What defines a well-documented method?** A common issue that users have raised is how to determine what a well-documented method is, especially if it is a method outside of their expertise. Here are a few tips to start thinking about whether or not a method is well documented and accepted:

- Do multiple studies use it?
- Do experts in that method or field consider it to be well documented and accepted?
- If it is a model, does the source report how the model was validated?

Tip 5: **Document your rationale:** At a minimum, document your rationale for the overall confidence level chosen. It is even more useful to document how a body of evidence scored for each criteria, with a brief explanation of why. This way, you can go back and understand how you graded something and your reasoning. This is especially helpful if multiple people are doing evidence assessments that will be compared, or if multiple people are doing evidence assessments for different linkages in a single results chain.

Understanding what the evidence is telling you:

Tip 1: **Existence vs. Predictability:** Assessing the available evidence for a particular hypothesis requires two considerations. The first is the existence of a relationship between two aspects of a system—does a change in one aspect lead to some change in the other? The second consideration, which is dependent on the first, is the predictability of that change. Do we have evidence to show how one aspect will change with the other? Often contextual information is required to understand the predictability of a relationship. It is important to note whether you are simply assessing strength of evidence for existence of a relationship, or whether you can also say something about the predictability of the expected change[4].

For example, a project team that we worked with was examining the evidence that connected a crop subsidy to regional stunting in children. Existing evidence for the relationship between subsidies and stunting was not sufficient to enable the team to predict the amount that stunting would be reduced, but new analysis by the project team revealed that adding annual regional precipitation as a control variable in the modeling of the relationship between the subsidy and stunting helped explain the relationship in a way so that predictions could be made about the effectiveness of the subsidy. While the existence of a relationship between a crop subsidy and stunting had been previously known, the extent of that change was now more predictable.

Tip 2: **The importance of 'other factors':** Strength of evidence also needs to take 'other factors' into consideration. Some fields call these other factors confounders, mediators, or effect modifiers. Broadly, however, it is important to consider that there might be other factors affecting the relationship between the two aspects of the system that are the focus of the hypothesis you are using the rubric to test.

Consider a hypothetical relationship between A (tidal marsh habitat area) and B (coastal flooding). A large body of evidence might describe the existence of the relationship between tidal marsh area and flooding, but other factors like sea-level rise might also influence coastal flooding (B). Those other factors might not be the primary driver you are considering, but they may be important in the estimation of coastal flooding. The existence of these other factors will likely lower the evidence grade between nodes A and B because they reduce the applicability and consistency of the evidence that links A and B. Alternatively, the strength of evidence for the influence of these other factors on the intervention effects can be directly considered (and these other factors could subsequently be added into a conceptual model, as illustrated in Figure 1)[5].

Tip 3: **Visualize the evidence assessment:** Use results chains, situational analysis or similar methods to visualize your evidence assessment[6]. Confidence levels can be added to each linkage in these conceptual models which can be an easy way to show what parts of your model you are most sure of, and what parts represent untested hypotheses (i.e. evidence gaps)

[4] This text is a modified excerpt from the Building Ecosystem Services Conceptual Models document
[5] This text is a modified excerpt from the Building Ecosystem Services Conceptual Models document

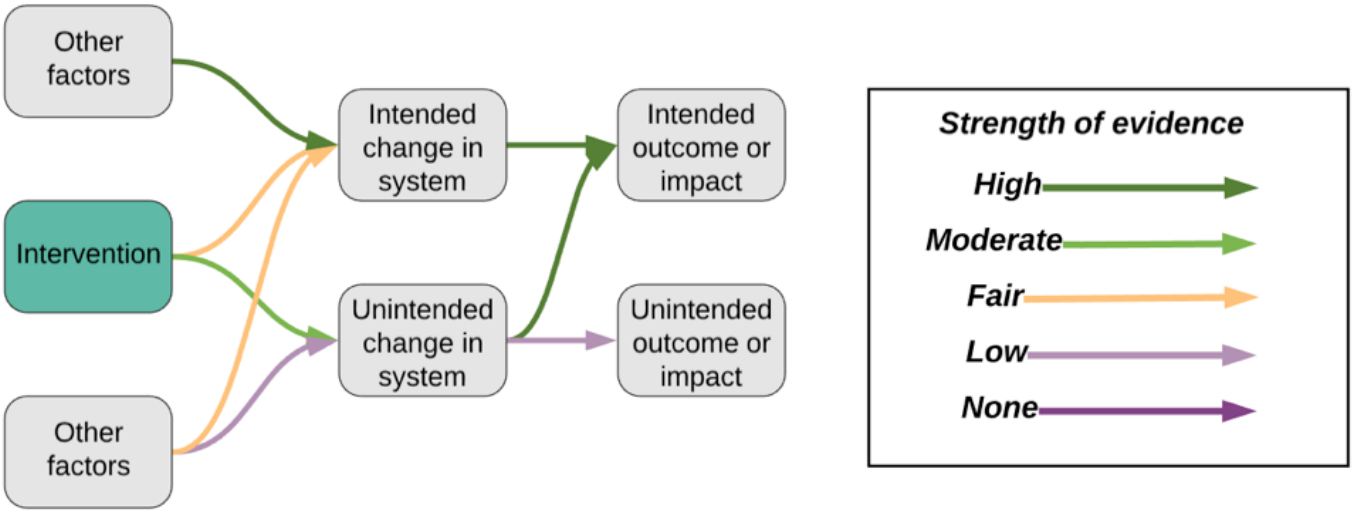

Figure 1. Example results chain with color-coded links that represent the strength of evidence available for each relationship in the diagram.

Identifying next steps:

What can you do with evidence scores once you have them?

Tip 1: **Prioritize research questions and monitoring resources:** Assigning evidence scores and understanding how they map to the system you are working in can help prioritize research questions. Are there certain important linkages or pathways that have low-scoring evidence? If so, research can be prioritized for those critical areas of the system where confidence is low. Monitoring resources can also be prioritized for those outcomes with lower scoring evidence pathways, in order to gather data on the parts of the system we know the least about.

Tip 2: **Communicate what is known about a system:** A conceptual model with evidence scores (as shown in Figure 1) is an easy way to communicate to stakeholders (the public, funders, policy-makers, etc.) what is known about a system and what isn't. It also can be a basis for communicating with researchers outside of your area of expertise, to begin discussions about cross-sector research planning.

Strength of evidence based on the rubric will be one of many factors that feed into decision-making processes. Strength of evidence scores need to be assessed in combination with effect size, political will, perceptions, cost and feasibility of action, and other factors. Within this context, having a common basis by which to assess and communicate insights about cross-sector evidence (using a rubric such as this one) can be valuable to support structured decision making.

Tip 3: **Understand risk better:** Looking at the confidence in a chain can inform risk assessment and decisions about whether or not and how to proceed with an intervention. For example, we may have very low confidence in a few links in a chain. If those links are high risk (because they are very important to a key stakeholder, or associated with high costs, etc.), we may not choose to proceed with an action. Alternatively, we may identify additional interventions to mitigate those risks, or invest in monitoring and evaluation to increase our confidence and understanding, and ability to adaptively manage the project[6].

[6] Modified excerpt from the Bridge Collaborative Practitioner's Guide

Tip 4:

Compare evidence strength to perceived level of evidence: Sometimes there is a mismatch between what the evidence indicates, and what key stakeholders in the system believe to be true – if your objective is to influence decisions or actions, resolving the gap between evidence and belief will be critical. Initiating conversations to navigate these differences may be important, as a first step. Asking questions such as "why do these differences exist?" and "what stories or events in the past have shaped this belief?" "what are the current incentives to maintaining this belief?" are good places to start.