Guidance on the HIP Criteria Tool

Michelle Weinberger

Karen Hardee

Annie Preaux

Maria Carrasco

February 15, 2023

Note this version of the guidance, developed under the auspices of the HIP Technical Advisory Group, relates to the September 2023 version of the HIP Criteria Tool.

Table of Contents

1. Introduction

The Family Planning High Impact Practice (HIP) Initiative uses five criteria in assessing practices for inclusion as HIPs. Figure 1 shows the definitions of the five criteria and the source for each.

Figure 1. Five Criteria for Assessing if High Impact Practices are Proven or Promising

Criteria	How defined for HIP review purpose	Source
Impact	Sufficient evidence of impact as per the HIP Evidence Scale (see Figure 4)	Based on summary of evidence included in the impact section of the HIP brief
Applicability, Reliability, Generalizability	Range of contexts or settings showing impact Broad evidence of impact from multiple contexts or settings (see Figure 4)	Based on summary of evidence included in the HIP brief
Scalability	Evidence of scale of the practice from impact being implemented at scale (not only from pilots) (see Figure 4)	Based on summary of evidence included in the HIP brief
Affordability	Qualitative rating based on what we know about cost and affordability. This is not the same as cost effectiveness	Experience/expert opinion
Sustainability	Based on HIP Sustainability paper (https://www.fphighimpactpractices.org/hip-sustainability-paper/	Experience/expert opinion

The HIP Initiative further categorizes Service Delivery and Social Behavior Change (SBC) high impact practices as proven or promising, defined as:

- **Proven:** Sufficient evidence exists to recommend widespread implementation, provided that there is careful monitoring of coverage, quality, and cost.
- **Promising:** Good evidence exists that these interventions can lead to impact; more research is needed to fully document implementation experience and impact. These interventions should be implemented widely, provided they are carried out in a research context and evaluated for both impact and process. https://www.fphighimpactpractices.org/hip-development/

A range of criteria are used to determine if a HIP is a proven or promising practice, including evidence of its impact; the applicability, replicability, and generalizability of the practice; and its scalability; affordability; and sustainability.

This guidance introduces, and explain how to use, the HIP Criteria Tool, an Excel-based tool that the HIP Initiative has custom built for use to characterize the evidence in HIP briefs related to the 5 HIP criteria, and to assess the evidence for determining both proven and promising practices.

This guidance is intended for TAG members who are determining proven vs. promising designation for service delivery and social behavior change HIPs, along with those who are completing the HIP Criteria Tool. Additionally, this guidance is available on the HIP website for full transparency of the process of reviewing the evidence in the impact section and for the other HIP Criteria.

2. What is the HIP Evidence Scale?

Determining the appropriateness of evidence and the strength of that evidence to inform policies and programming is challenging. In developing the HIP Evidence Scale, embedded in the HIP Criteria Tool, the HIP Initiative sought to reinforce that HIP briefs are not intended to be systematic reviews; rather they are intended to accommodate a range of programmatic evidence. Given the range of interventions related to family planning, no one study type is considered best for measuring outcomes of all of the interventions (Figure 2).

Figure 2. HIP Outcomes

	Proximal Outcomes	**Related/Additional Outcomes**			
Family Planning Outcomes	Increase CPR, mCPR, birth spacing, decrease unwanted pregnancies, delay marriage/sexual debut (for adolescents)	Expand method choice, quality, and coverage	Reach diverse underserved groups	Address social and cultural barriers	Reduce financial barriers

The HIP Evidence Scale, shown in Figure 3, is used to assess the evidence in the impact section of HIP Briefs. The HIP Evidence Scale contributes to determining if the HIP is designated as proven or promising, along with the other criteria shown in Figure 1, and the reasoned judgment of the TAG following discussion of the practice and available evidence. The HIP Evidence Scale is adapted from a methodology for classifying evidence for public health, based on a typology by Gray (1997: 61; 2009).[1] The 'Gray' scale uses five levels/types to assess the strength of studies associated with a body of

[1]For more detail, see Hardee et al., a forthcoming paper on the development and use of the HIP Evidence Scale within the context of the HIP TAG and partnership.

evidence,[2] In keeping with the Gray scale, the HIP Evidence Scale also uses five levels of evidence, levels I through V. The first three levels of evidence (I, II, and IIIa) are all based on studies that include a control group or other rigorous design, with levels I and II also reflecting random assignment to intervention or control groups. Another grouping (levels IIIb, IV, and V) reflect evidence without a control group, including pre-post study designs without a control (IIIb); routine or program data (IV), and qualitative (V).

Figure 3. HIP Evidence Scale	
Level	**Type of Study**
Evidence with a control group	
I	Systematic review of randomized control trials (RCT)
II	Randomized control trials
IIIa	Control with pre/post design (non-randomized/quasi-experimental)
	Control with post-only design (non-randomized)
	Other rigorous design (e.g., propensity score matching)
	Systematic review of non-RCTs (quantitative)
Evidence without a control group	
IIIb	Pre/post design, no control
IV	Routine/program data (e.g., service statistics or other M&E data)
V	Qualitative
	Systematic review of non-RCTs (qualitative)
n/a	Other/unsure*
	*This designation is if one reviewer is unsure of the type of study and wants to flag it for another reviewer.

[2] For more detail about development of the Gray Scale see http://www.whatworksforwomen.org/pages/methodology.

Classifying the evidence in the impact section in each HIP brief provides a summary not only of the types of evidence available to assess each practice, the outcomes measured in each study, and whether they are positive and statistically significant or not, whether they show mixed results, or negative results. Within a tab of the HIP Criteria Tool, these parameters of the evidence can be displayed for ease of review. The spreadsheet, explained in more detail below, also summarizes the evidence of applicability, generalizability, and replicability, and the evidence of scalability. Figure 4 shows the conditions for each of the HIP criteria (column 1) corresponding to a proven (column 2) and promising (column 3) designation. The HIP TAG approved these tips for determining proven vs. promising designation for service delivery and SBC HIPs at its June 2023 meeting.

Figure 4 provides tips for determining proven or promising

Figure 4. Tips for determining proven/promising designation for HIPs using the 5 HIP Criteria

(1)	(2)	(3)
HIP Criteria	**Proven**	**Promising**
Impact	At least 4 studies with positive evidence at level I, II, or IIIa on the HIP Evidence Scale (with at least 3 studies with statistically significant results), with explanation for exceptions	At least one study at levels I, II and IIIA and/or at least 4 studies at levels IIIb, IV or V, with explanation for exceptions. *
Applicability, reliability, **generalizability**	At least 4 countries across more than one region	Fewer than 4 countries or evidence from only one region
Scalability	Broad evidence of implementation at reasonable scale for the HIP (at least 50% of studies implemented at a reasonable scale)	Evidence largely from pilots and/or small scale implementation (greater than 50% of the studies show implementation from pilots and/or small scale implementation
Affordability	Not included in determining proven/promising designation given paucity of evidence on costs. Authors of HIP Briefs encouraged to include existing evidence of affordability	
Sustainability	Not included in determining proven/promising designation. Authors of HIP Briefs encouraged to review the sustainability checklist in the White Paper and to include evidence of sustainability.	

*Note: The TAG should be notified if the practice does not meet the threshold of promising.

The HIP TAG uses the HIP Evidence Scale as input to reviewing the five criteria for determining if a HIP is proven or promising. The TAG reviews the summary statistics generated from the HIP Evidence Scale in the HIP Criteria Tool. The HIP TAG incorporates the strength of evidence related to the practice into their deliberations in determining which category (proven or promising) a HIP Brief falls into. When assessing the evidence of impact, the TAG seeks to determine if the body of evidence included in the impact section is based on the most appropriate study types for the outcome addressed by the HIP.

3. Extracting Information for the HIP Criteria Tool and Summarizing the Results

For each service delivery or SBC HIP Brief being developed, the HIP initiative will designate someone to complete the HIP Criteria Tool, including the HIP Evidence Scale.

HIP Literature Review. Several pieces of information are needed to complete the HIP Evidence Scale component of the HIP Criteria Tool. Extracting this information begins with the initial literature review and extraction of background information that is put into a HIP Literature Review Spreadsheet (this is a separate spreadsheet from the HIP Criteria Tool). Information needed for the HIP Evidence Scale on the studies included in the impact section of the HIP brief is included in several columns of the literature review spreadsheet. In addition to background information on the study (e.g., the reference, the intervention, the results), the information needed for the HIP Evidence Scale include columns, some with pull down menus:

- **Type of study** (selected from the HIP Evidence Scale study types in Figure 3)
- **Specific outcome measured** (the primary outcome to assess should be the impact of the interventions on mCPR/contraceptive use, but for some interventions additional outcomes (e.g., communication, norms) may be appropriate; the evidence for each outcome should be assessed in separate Excel files. Each file will focus on a broad outcome (e.g., mCPR, attitudes) but this can capture differences in measurement of the outcome across studies (e.g., used method at last sex, client left with a method).
- **Impact findings significant?** (significant, not significant, no test for significance, mixed)
- **Direction of finding** (positive, negative, no difference, mixed)

Additionally, information to assess generalizability and scalability are also extracted:

- **Study focused on specific sub-population** (e.g., sex workers) (yes, no)
- **Study focused on specific context** (e.g., refugee camp) (yes, no)

- **Scale of intervention studied** (pilot, implemented at small scale, implemented at reasonable scale)

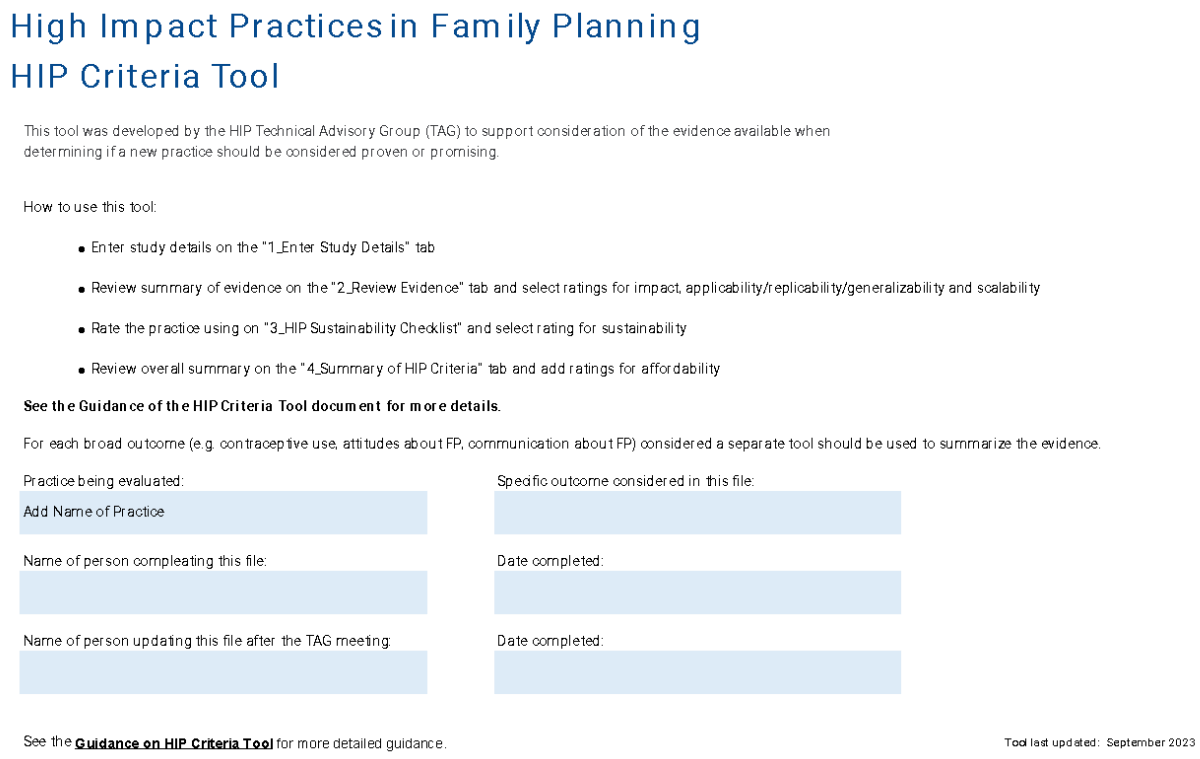
In the course of developing the HIP brief, the expert group working on it might identify additional studies to include as evidence in the brief, including in the impact section. Information from the additional studies should be added to the same literature review spreadsheet.

For service delivery and SBC briefs, relevant information on the studies used in the impact section of the HIP Briefs should be transferred to the HIP Criteria Tool Excel file.

HIP Criteria Tool Introduction Tab

An introductory tab in the HIP Criteria Tool provides information on how to use the tool and asks for the person filing in the tool to indicate:

- The practice being assessed
- Specific outcome considered in the spreadsheet
- Prior to TAG review
    - Name of the person completing the spreadsheet
    - Date of completion of the spreadsheet
- Following TAG review
    - Name of the person completing the spreadsheet
    - Date of completion of the spreadsheet



High Impact Practices in Family Planning

HIP Criteria Tool

This tool was developed by the HIP Technical Advisory Group (TAG) to support consideration of the evidence available when determining if a new practice should be considered proven or promising.

How to use this tool:

- Enter study details on the "1_Enter Study Details" tab
- Review summary of evidence on the "2_Review Evidence" tab and select ratings for impact, applicability/replicability/generalizability and scalability
- Rate the practice using on "3_HIP Sustainability Checklist" and select rating for sustainability
- Review overall summary on the "4_Summary of HIP Criteria" tab and add ratings for affordability

See the Guidance of the HIP Criteria Tool document for more details.

For each broad outcome (e.g. contraceptive use, attitudes about FP, communication about FP) considered a separate tool should be used to summarize the evidence.

Practice being evaluated: Add Name of Practice

Specific outcome considered in this file:

Name of person compleating this file:

Date completed:

Name of person updating this file after the TAG meeting:

Date completed:

See the **Guidance on HIP Criteria Tool** for more detailed guidance.

Tool last updated: September 2023

Following the introduction, this Excel file has three main tabs:

Tab 1: Detailed evidence review

Tab 2: Summary of evidence

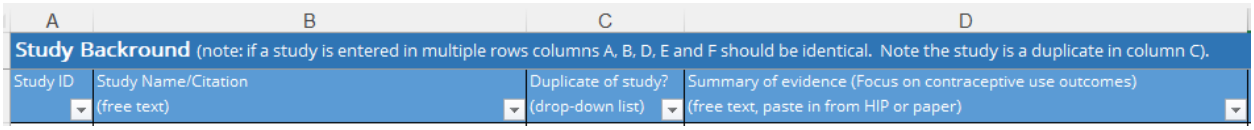
Tab 3: Summary of HIP Criteria

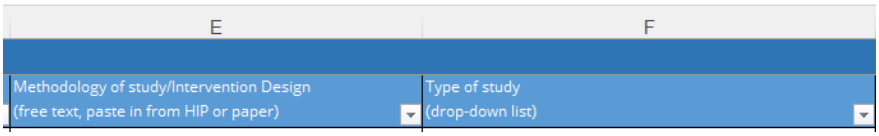
Two additional tabs are used for automatic calculations to populate Tabs 2 and 3.

Tab 1. Detailed evidence review

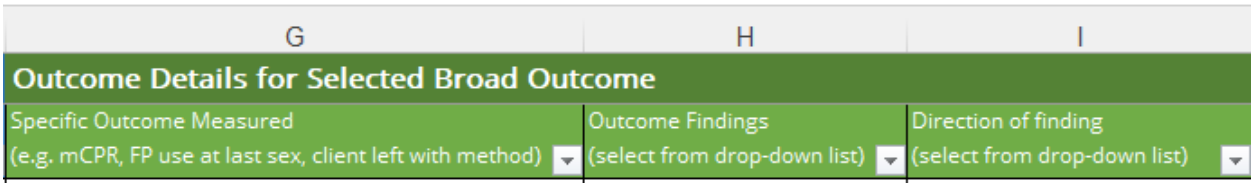
First, the details from each study are summarized in the 18 columns in Tab 1, shown below:

General Study Background should be entered in columns A to F. If a study is entered in multiple rows (see more below) columns A, B, D, E and F should be identical. Note the study is a duplicate in column C

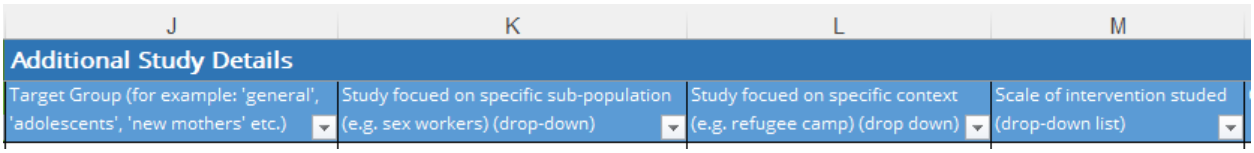


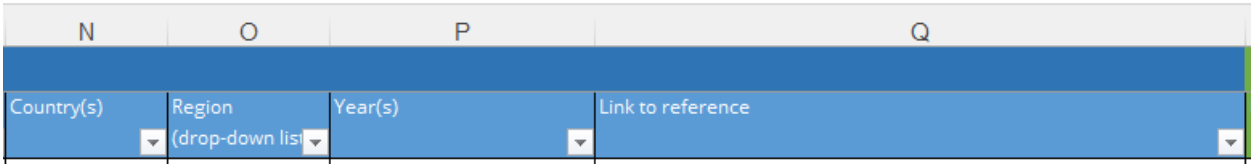


Outcome Details for the Selected Broad Outcome should be entered into columns G to I. Different HIP Criteria Tools should be used for each broad outcome (e.g., contraceptive use, FP attitudes, FP communication) considered in cases where the HIP warrants evaluating multiple outcome areas. b

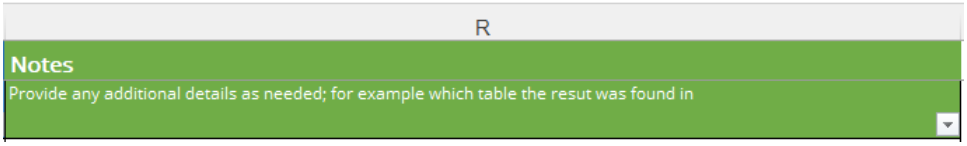


Additional Study Details should be entered into columns J to Q.





Finally, column R can be used to provide any additional notes, for example to indicate which table the outcome result was taken from.



The information for Tab 1 comes from the literature review excel file, and as needed, from additional studies found by the expert group, if those are put into a separate file.

In some cases, a single study will be entered in multiple rows in the Excel file:

- For studies with more than one outcome that measures the same concept (e.g., the study might report contraceptive use based on both used method at last sex and current use of modern method), select the outcome that most closely aligns with the HIP. If the outcomes are similar (e.g., different ways to measure contraceptive use), the most relevant one should be selected.
- For studies with more than one intervention or intervention group and results are reported separately, these should be entered as separate rows (with details provided in the "specific outcome measured"). Examples include intervention A vs intervention B, region 1 vs region 2. Some studies have multiple different intervention arms -- these should be on separate rows in the Excel file. Likewise, some studies analyze the impact of different intervention components – these should also be on separate rows. Reviewers should use their judgment to potentially limit the number of results that are included for studies that tested many small differences to an intervention (e.g., in this case might only want to include 1 or 2 key models). Column C (duplicate of study) should be used to indicate multiple results from the same study in order to avoid double counting the number of studies.
- For studies that report outcomes for different demographic groups (e.g., unmarried women vs married women), these should be reported as separate rows (with details provided in the "specific outcome measured"). Reviewers should use their judgment to potentially limit the number of results that are included for studies that report outcomes for many different groups (e.g., in this case might only want to include 1 or 2 key groups). Column C (duplicate of study) should be used to indicate multiple results from the same study.
- For interventions/studies that are included in more than one article, multiple articles should only be included if reporting on different results following the criteria above (e.g., the articles address different outcomes from the same study). Even though the results are in a different article, the study should still be indicated as a duplicate in column C (duplicate of study), so that study counts are not inflated in the count of the number of studies.
- For studies undertaken in multiple countries, add a row for each country if the results are given per country. Column C (duplicate of study) should be used to indicate that the results from the same study come from multiple countries in order to avoid double counting the number of studies.

Tab 2: Summary of evidence

Tab 2 also contains automatically generated statistics related to three of the five criteria for determining if a HIP is proven or promising. The first criterion is impact.

Impact Summary: HIP Evidence Scale

Based on the information in Tab 1, the impact summary table in Tab 2 will be automatically filled in. This table shows the types of studies included in the impact section and the strength of the evidence for the practice. Note that the information in the tables in this guidance document are for illustrative purposes only. The impact summary shows the number of results and the number of studies included in the impact section, and the number/percentage of those results. In this example, there are 25 results (and 21 studies).

The studies in the impact summary are grouped by studies with a control group (levels I, II and IIIa on the HIP Evidence Scale) and studies without a control group (levels IIIb, IV and V). The first two results columns indicate studies with positive and statistically significant results, and positive results without a test of significance of the results. The distribution of results (in %) for these two columns is indicated in green. The next two columns indicate nonsignificant results and results where there was no difference, with the distribution of results indicated in yellow. The firth column indicates studies with negative results; its distribution with the green and yellow categories is shown in red. This example shows that 11 of the 25 results are positive with significant results (with 9 of those results coming from level II (RCT) and level IIIa (control with pre/post tests and other rigorous design).

Impact Summary using the HIP Evidence Scale for: Add Name of Practice

Level	Type of study	# positive significant results	# positive results but no significant test	# non-significant results	# no difference results*	# negative results*	Other (inc. mixed results)	Total # Results*	Total # Unique Studies*
Studies with a Control									
I	Systematic Review of RCT	0	0	0	0	0	0	*0*	*0*
II	RCT	**3**	0	**1**	**1**	0	0	*5*	*4*
IIIa	Control with pre/post (non randomized/quasi-experimental)	**4**	**3**	**1**	0	0	**1**	*9*	*8*
	Control with post only (not randomized)	0	0	0	0	0	0	*0*	*0*
	Other Rigorous Design (e.g. propensity score matching)	**2**	**1**	**1**	0	**1**	0	*5*	*4*
	Systematic Review of non-RCTs (quantitative)	0	0	0	0	0	0	*0*	*0*
Studies without a Control									
IIIb	Pre/post no control	0	0	**1**	**2**	0	0	*3*	*2*
IV	Routine/program data	0	0	0	0	0	0	*0*	*0*
	Other non-rigorous design	**2**	0	0	0	0	0	*2*	*2*
V	Qualitative	0	0	0	0	0	**1**	*1*	*1*
	Systematic Review of non-RCTs (qualitative)	0	0	0	0	0	0	*0*	*0*
n/a	Other/unsure	0	0	0	0	0	0	*0*	*0*
	Total Results*	**11**	**4**	**4**	**3**	**1**	**2**	***25***	***21***
	Distribution of studies by result	44%	16%	16%	12%	4%	*8%*		
		60%		28%					

*In some cases one study may contribute multiple results; this table shows the total number of outcomes included not unique studies

†Includes studies where this result was significant, or no significance test was conducted

Select the rating based on the HIP evidence summary and tips for determining proven/promising

Agreed Impact Score	Proven

If an exception was made to the proven/promising designation please explain below:

HIP criteria met; most evidence shows positive results.

Tips for determining proven/promising designation:

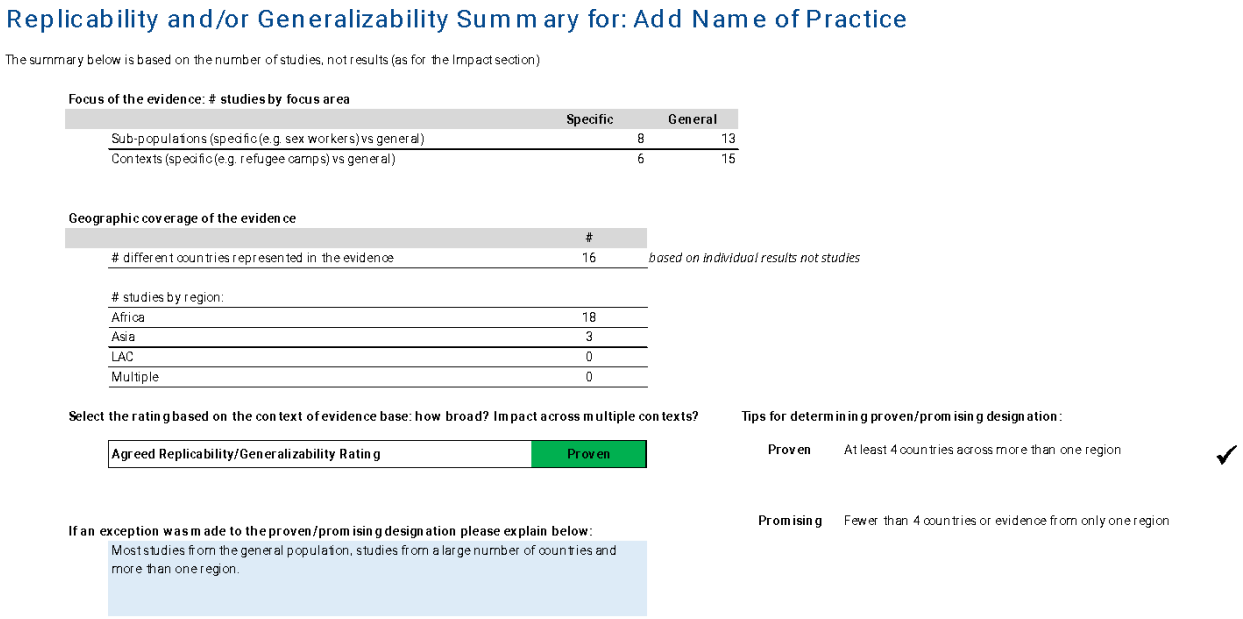
Proven	At least 4 studies with positive evidence at level I, II, or IIIa on the HIP Evidence Scale (with at least 3 studies with statistically significant results), with explanation for exceptions	✓
Promising	At least one study at levels I, II and IIIa and/or at least 4 studies at levels IIIb, IV or V, with explanation for exceptions	

Based on the Impact Summary, an "agreed impact score" for the evidence is selected: proven, promising, or no evidence (indicating that there is insufficient evidence to consider the practice a HIP). The TAG also provides an explanation for why they selected this score. This score is determined by

reviewing the number of results and studies in the impact section, the percentage of those results that are positive with statistical significance and positive but with no test of significance, etc. In this example, given the number of results and studies and the percentage of results coming from positive and statistically significant results (56%) with an additional 4% of results being positive without a significance test, the HIP TAG agreed to a "Green" impact score. This reasoning will be added as an explanation in the blue shaded box.

Replicability and/or Generalizability Summary

The second criterion for which summary statistics are tallied in Tab 2 is replicability and/or generalizability. The tallies for the dimensions of 'focus of the evidence' and 'geographic coverage of the evidence' are based on number of studies rather than number of results.



Replicability and/or Generalizability Summary for: Add Name of Practice

The summary below is based on the number of studies, not results (as for the Impact section)

Focus of the evidence: # studies by focus area

	Specific	General
Sub-populations (specific (e.g. sex workers) vs general)	8	13
Contexts (specific (e.g. refugee camps) vs general)	6	15

Geographic coverage of the evidence

	#	
# different countries represented in the evidence	16	*based on individual results not studies*
# studies by region:		
Africa	18	
Asia	3	
LAC	0	
Multiple	0	

Select the rating based on the context of evidence base: how broad? Impact across multiple contexts?

Agreed Replicability/Generalizability Rating	Proven

Tips for determining proven/promising designation:

Proven	At least 4 countries across more than one region	✓
Promising	Fewer than 4 countries or evidence from only one region	

If an exception was made to the proven/promising designation please explain below:

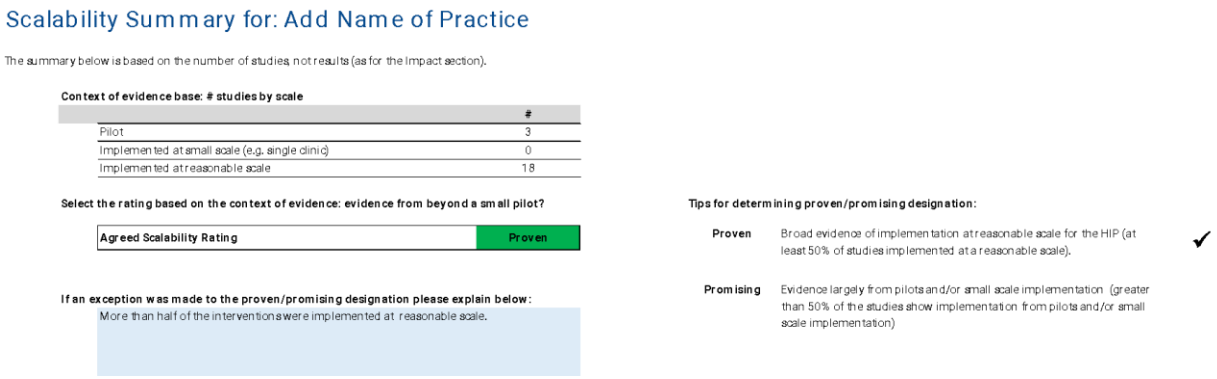
Most studies from the general population, studies from a large number of countries and more than one region.

As with the impact summary, the HIP TAG reviews the distribution of the studies by these dimensions (breadth of the evidence base – focused or general; and geographic coverage of the evidence) to agree on a replicability/generalizability rating.

In this example, because the evidence came from more than one region (Africa and Asia) and because the evidence includes a range of general (13) and specific (8) populations, and a range of specific (6) and general (15) contexts, the HIP TAG selected a 'Green' rating. This explanation should be typed into the blue shaded box to explain why the score was selected.

Scalability Summary

The third criterion for which summary statistics are tallied in Tab 2 is scalability, with numbers of studies that reflect 'pilots', 'implemented at small scale', and 'implemented at reasonable scale'. In this example, 18 of the 21 studies were of interventions implemented at reasonable scale, earning a scalability rating of 'Green', with the explanation added to the blue shaded box.



Scalability Summary for: Add Name of Practice

The summary below is based on the number of studies, not results (as for the Impact section).

Context of evidence base: # studies by scale

	#
Pilot	3
Implemented at small scale (e.g. single clinic)	0
Implemented at reasonable scale	18

Select the rating based on the context of evidence: evidence from beyond a small pilot?

Agreed Scalability Rating	Proven

If an exception was made to the proven/promising designation please explain below:

More than half of the interventions were implemented at reasonable scale.

Tips for determining proven/promising designation:

Proven	Broad evidence of implementation at reasonable scale for the HIP (at least 50% of studies implemented at a reasonable scale). ✓
Promising	Evidence largely from pilots and/or small scale implementation (greater than 50% of the studies show implementation from pilots and/or small scale implementation)

Tab 3: HIP Sustainability Checklist

Tab 3 contains a checklist to help determine potential sustainability of the HIP. Although sustainability does not have an explicit proven or promising designation, those compiling the HIP Criteria Tool can suggest a rating on sustainability that can be considered by the HIP TAG in its discussions about the practice.

HIP Sustainability Checklist for: Add Name of Practice

For the HIPS sustainability from the perspective of scale-up, capturing four dimensions:

1. Political and policy sustainability
2. Financial sustainability
3. Organizational sustainability
4. Maintaining equity of access, coverage and quality of services

See more in the HIP Sustainability Paper: https://www.fphighimpactpractices.org/hip-sustainability-paper/

Using the list of questions below, please make a determination about this HIPs potential for sustainability.

Agreed Scalability Rating	Sustainability not well addressed

Explain why this score was selected

Limited evidence of operation under routine settings

Questions and Considerations About Sustainability in Relation to the HIP Initiative, for Authors, Reviewers, and Implementers

Questions to ask about sustainability of the HIPs	Is this addressed in the HIP? How?
Is the practice clearly defined?	Yes
Is the practice programmatically relevant?	Yes
Has the practice been scaled beyond its initial setting to reach a larger portion of the target population?	Partially
Has the practice been tested under routine operating conditions and existing resource constraints of relevant health systems (e.g. beyond controlled pilot settings)?	Partially
Is the practice simple enough (e.g. it is not overly complex or reliant on charismatic individuals) such that the inputs required to implement it are manageable at large scale?	No
Additional questions to be considered by implementers	
Is there an organizational home and leadership for the practice? If the practice crosses organizations/ministries, are they aligned on the practice implementation and are the roles and responsibilities for implementation clear? Are all relevant stakeholders involved in the process of scaling up the HIP?	
Have the stakeholders reached consensus on their expectations for scale up?	
Is the source of funding for the practice scaled up reliable over a reasonable time period? What are the sources of funding (national/donor)?	
Are policies, regulations, and other system components in place for sustainable scale up of the practice?	
Does the definition of sustainability of the practice include a focus on equity and quality— not just financial sustainability?	
Is there a plan to implement M&E and learning as part of the scale-up process and/or the scaled-up practice to assess sustainable implementation of the practice?	

Tab 4: Summary of HIP Criteria

The fourth tab summarizes the ratings for the three criteria – impact; replicability and/or generalizability; and scalability. Unless the TAG makes an exception with an explanation for the rationale for the exception, to be proven, a practice should show proven impact and proven for at least one of the other two criteria. In the example in this guidance document, the three criteria were all proven.

Space is also provided to add ratings and explanations for the other two criteria: affordability and sustainability, however these criteria do not have explicit proven or promising designations although they can factor into the TAG's discussion on a practice. In reviewing the three scores, if there is a mix of ratings across the criteria, for example if the scores are proven for impact and promising for the other

two criteria, or are promising for the other two criteria, the TAG will need to make a decision anchored on impact.

Summary of HIP Criteria for: Add Name of Practice

Ratings and notes for the first three HIP Criteria are automatically populated from the information entered on the previous tab (to revise these please return to the previous tab).
Ratings and notes should be added for the affordability and sustainability HIP Criteria below.

Criteria	How defined for HIP Review purpose	Source	Rating	Documentation of exceptions to criteria
Impact	Sufficient evidence of impact as per the HIP Evidence Scale	Based on the HIP Evidence Scale *(see tab 2)*	Proven	HIP criteria met; most evidence shows positive results.
Applicability, Reliability, Generalizability	Range of contexts or settings showing impact. Broad evidence of impact from multiple contexts or settings	Based on a summary of evidence included in HIP Evidence Scale *(see tab 2)*	Proven	Most studies from the general population, studies from a large number of countries and more than one region.
Scalability	Evidence of scale of the practice from impact being implemented at scale (not only from pilots)	Based on a summary of evidence included in HIP Evidence Scale *(see tab 2)*	Proven	More than half of the interventions were implemented at reasonable scale.
Affordability	Qualitative rating based on what we know about cost and affordability. This is not the same as cost effectiveness	Experience/expert opinion	Not included in determining proven/promising designation given paucity of evidence on costs. Authors of HIP Briefs encouraged to include existing evidence of affordability	
Sustainability	Based on HIP Sustainability paper (https://www.fphighimpactpractices.org/hip-sustainability-paper/)	Experience/expert opinion *(see tab 3)*	Not included in determining proven/promising designation. Authors of HIP Briefs encouraged to review the sustainability checklist in the White Paper and to include evidence of sustainability.	

4. Presenting the HIP Evidence Scale Summary to the HIP TAG to make a final determination

Once the HIP Evidence Scale Excel File has been completed, it is presented to the HIP TAG to make a final determination about the proven or promising designation for the HIP brief. Tab 4 of the HIP Criteria Tool has space for the final determination of the HIP TAG to be recorded.

Final TAG Determination for the practice

For a HIP to be classified as proven, a practices should show proven impact and proven for at least one of the other 2 criteria. Any exceptions should be documented below.

Based on the summary above and TAG discussion, the TAG has agreed to rate this practice as: **Proven**

Summary of TAG discussion on rating

The presentation of the evidence for each HIP using agreed proven and promising criteria will be available in TAG meeting reports moving forward, along with a summary of the discussion of the TAG and their final determination for each HIP Brief.

The HIP Initiative maintains internally the HIP Criteria Tool for each relevant (e.g., service delivery and SBC) HIP Brief. The information is available on request.