Health Systems Strengthening Evaluation Collaborative

Global Stakeholder Analysis

July 2022

Natasha Palmer, Veena Sriram, Sara Bennett

Table of Contents

1. Introduction

Health systems strengthening (HSS) is widely understood to be key to achieving universal health coverage and to ensuring robust responses to health emergencies. In recent decades, global health investors have put more attention and investment towards HSS, leading to accelerated efforts to evaluate HSS policies and programs initiated by those investments. Yet, a common definition and framework for how to evaluate HSS interventions remains elusive, hampering efforts to strengthen, coordinate and amplify HSS programs.

The Health Systems Strengthening Evaluation Collaborative (HSSEC) brings together key global and national stakeholders to suggest ways to strengthen the quality of evaluations of health systems strengthening (HSS) investments in LMICs and to improve coordination across stakeholders in this space. The Collaborative believes that to move HSS evaluation beyond its current fragmented form, leadership and commitment for advancing and changing ways of working must come at least partially from the joint action of three key groups of stakeholders: (i) country-level stakeholders, including governments, practitioners, and communities, (ii) donors that fund HSS and HSS evaluation, and (iii) evaluators and academics who are involved in HSS evaluation.

Methodology

This study was designed as a prospective stakeholder analysis, utilizing qualitative research methods, such as semi-structured interviews and document review.

To select informants involved in HSS investment processes at the global-level, the research team relied on purposive sampling and snowballing techniques. The team conducted 23 interviews with 25 respondents (two interviews were carried out together with two respondents) via Zoom and Microsoft Teams. Verbal consent to participate in the research was provided to our team at the beginning of each interview and the questions were shared with the interviewees in advance. Qualitative analysis techniques such as pattern spotting were applied to interview transcripts and these insights were triangulated with document review data.

The interim findings were shared amongst the research team and with the HSSEC Executive Committee and the feedback provided was used to strengthen the findings presented in this final report.

See more details on the methodology in Annex 4.1.

2. Summary of key findings

2.1 Shared understandings and definitions of health systems strengthening remain a challenge amongst global health investors

There is growing recognition among global health funders of the importance of health systems for achieving global health goals. This was driven by a recognition of the need for HSS due to recent emergencies, such as Ebola and COVID-19, and from decades of learnings around implementation successes and challenges.

The growing focus on HSS however has not been accompanied by cohesion around a common understanding of what constitutes HSS.

Respondents frequently pointed to ambiguity between health systems support and HSS.

"…so the very kind of notion of what is health system strengthening I think has become very distorted in [global health initiative]. And what's counted as health system strengthening is largely support." IDI10, philanthropic organization

"…health system strengthening in some ways can be described as everything we do. But it also can be limited, quite limited to a very small sub sector. So depending on how you define it, the relative importance can really vary." IDI11, bilateral agency

There were differences in how global health investors conceptualized HSS that went beyond the support vs strengthening debate.

For some respondents, HSS was a distinct set of investments in support of, or in addition to, disease-oriented program investments (i.e., cold chain investments). For others, HSS was viewed as integral to their funding approach, and was a value or guiding principle, rather than a discrete investment. HSS was viewed by these investors as cross-cutting all investment decisions, including disease-specific goals.

"We really feel the health system strengthening aspect is that top priority, and it's a prerequisite to reach those disease goals and other health goals in reproductive, maternal child health, etc." IDI12, bilateral agency

This **lack of shared understanding and definitions** around HSS leads to a confused discourse around evidence, what it is and what it is needed for.

"People just come out in their camps of you should do strengthening over support, you shouldn't invest in the verticals, you should just invest in the system strengthening, and that's your answer…But that that bit is not based really on looking at the evidence that…" IDI19, multilateral donor

2.2 Types of evidence global health investors are seeking in HSS evaluations

There was divergence in our interviews on the *type* of HSS evaluation evidence sought or desired, and in the *end purpose* for gathering that evidence. We were able to discern three categories of "evidence need" that came out of the interviews « Return on Investment », « Relative Effectiveness of HSS Investments » and « Implementation Research » (Figure 1).

Figure 1 : Types of evidence sought by global health investors

Return on Investment	Relative effectiveness of HSS investments	Implementation, policy analysis and other forms of systems research
- Link between HSS investments and disease outcomes/impact - Attribution or contribution of investments to impact - Cost effectiveness - Time to impact	• Resource allocation within HSS • Cost effectiveness • Funding utilization/absorbtion	• Implementation research ("how" questions) • Political economy analysis • Case study research • Tacit knowledge

Return on Investment

Many respondents stressed the importance of measurable results and « Return on Investment », particularly through improvements in health outcomes. Respondents discussed needing to have clear evidence of impact in terms of "lives saved" that could be tracked within immediate budgeting cycles, establishing a "line of sight" between investments, outputs and health outcomes

Bilateral donors expressed the need for such evidence to satisfy lawmakers' (and by extension, voters') desire to see results from resources invested. Other donors indicated that such evidence was needed to either redirect or increase funds for HSS investments.

"…to actually be able to understand the correlation between investments in health systems and health system strengthening to disease outcomes. And that's the bit I'm really keen to see. I think that's the area that the world has struggled forever." IDI19, Philanthropic organization

Despite a number of global agreements[1] committing funders to seek to evaluate the contribution of their investments to collective outcomes rather than attributing specific outcomes to their investments alone, respondents noted that funders were still often interested in attribution. Other respondents acknowledged this impulse, but noted the methodological challenges in securing such evidence.

"And just even simple things, like our impact…impact is what they [donors] all want…Because they want to look at value for money. But we don't have systematic methods for how to do that. And we are all contributing to different things. And so then you can imagine for a donor, they want to be able to add up…I put money in GAVI, GFF, Global Fund, what was our impact for the money invested? Well, we have double counting across some of that. And we have different methods. Some people are more robust or more transparent than others." IDI2, multilateral organization

Return on Investment decisions were also perceived to be closely linked to "tradeoffs" in decision-making around global health investments. Underlying these concerns was the sense that funding for global health was finite, and that investors would need to balance

[1] See for example "The Three Ones in Action" UNAIDS, 2005: https://data.unaids.org/publications/irc-pub06/jc935-3onesinaction_en.pdf

multiple dimensions in their decision-making: impact of investment on health outcomes, time to impact (lives saved in the near-term v. the long-term), perceptions around the disease- or health-conditions targeted, and national interests on the part of bilateral donors. The role of evidence was raised by some respondents as important in understanding the consequences of these trade-offs.

"And it's that that's what I would love to see and have, because…it fundamentally gets back …to a trade off with limited resources. And so when you were only given a choice of saving a life now, versus, we are willing to trade off lives now for a stronger health system later 10 years' time. And that just doesn't wash in a global decision making community, especially when you have communities at the table. And so that's the bit…the nuance of how do you measure? How do you combine the two of measuring, strengthening and how it has an impact in on the current situation… it's really understanding where your trade off really is."
IDI19, Philanthropic organization

"there's this perception of a tradeoff, right? So I would have to sacrifice results to strengthen health systems. And I think that's the piece that needs to be broken down where you can say, you don't have to sacrifice results, and you can strengthen health systems. And I don't think there's any clear evidence of that. And so that's, that's the sort of where you get when push comes to shove on that it's like, well, even if I believe strengthening health systems is important. I can't sacrifice the lives saved now piece of it." IDI5, bilateral agency

Evidence on relative effectiveness of HSS investments:

Several respondents indicated that evidence on the relative effectiveness of HSS investments (i.e., the relative impact of one type of HSS investment over another) was important to decision-making. However, similar to their perceptions of the challenges with much of the "Return on Investment" evidence, respondents indicated that such relative measures of effectiveness, such as cost effectiveness or evidence linking macro-level systems investments to performance, were not widely available.

"For every country, just give me a league table that shows me the relative cost effectiveness of different sorts of health systems strengthening interventions so that I can just say, Okay, well wait a minute, I should be starting with I should start with the HMIS, invest this much before I move down to investing this much in cold chain before I move this… something that really synthesizes and pulls together?" IDI9, multilateral agency

Implementation research, policy analysis and other forms of systems research:

Fewer respondents noted the importance of other forms of research – such as implementation research, policy analysis and other types of systems research – in driving decision making. Some respondents noted that this type of research was needed in combination with the aforementioned research on impact and effectiveness. Some respondents who supported this approach noted the challenge of primarily relying on standardized evaluation approaches to highly contextual programs and policies pertaining to HSS, noting that implementation or operational research that delved into the mechanics of HSS programs and policies would provide richer analysis of barriers and facilitators to program success.

"You don't need just one type of evidence analysis, you need the policy analysis, you need the application, so that you need the metrics, you need all of those things to strengthen health systems, and not just indicators of health outcomes, which is just one very small and often poorly linked to health system strengthening efforts." IDI1, multilateral agency

"And the economist perspective dominated the evaluation approach to PBF (performance-based financing) in the last seven or eight years. At the expense of what many of us were arguing for, which is more of the operational implementation research at the local level. Now, I'm hoping that because of the limited utility of some of these RCTs, the wheels will swing back more towards implementation, operational research…" IDI6, multilateral agency

3.3 How is HSS evidence communicated and framed to global health investors and other audiences ?

Several respondents raised the concern that the health systems research community has not been as successful as some other "disease-specific" groups in making the case for investment in their field.

"the strength of the disease specific voices is incredibly strong. And there's no way to counter that voice without the evidence. Where's the evidence, if we take money from the recommendation of antiretroviral treatment to system strengthening, that it will pay off? And there isn't any, to be honest. " IDI10, philanthropic organization

"So the rest of the health systems research community hasn't been able to strengthen the case enough for the investment in the types of research [needed to provide evidence on HSS investments]. And I think the way to do that is partly scientific, by improving the communication, or around that science, and how it how it gets communicated to answer the questions that are being asked" IDI1, multilateral agency

Other respondents believed that the health policy and systems research community could go further in identifying targeted and tangible ways in which HSS research findings are directly applicable to global health funding patterns.

"Where it gets hurt is the lack of specificity, and this issue of absorption. And so, evidence does play quite a bit of role. I mean, there's lot of talk around HRH, human resource shortfall training, supportive supervision, quality of care. But how that evidence gets used is hard to put a finger on, to be honest with you. Because this funding is allocated in these broad generic parts. It's not really clear what the intervention is." IDI22, multilateral agency

Related to the issue of framing and communication were global health investors' disparate conceptualizations of HSS, which led to varied interpretation of evidence. As a result, a few respondents noted the 'art appreciation' involved with interpreting HSS evidence.

"I think the issue with evidence, and especially if you think about health systems and UHC is that the interpretation of evidence, I would argue is more of an art than a science. And so, really having and again, it comes back to having a really clear conceptual grasp of what is meant by these terms, how to interpret them." IDI7, multilateral agency

"most of the evidence is highly academic, it's very, very gray, it's not easy to read, was it a success or not? How much what was the cost effective? So, it's more art appreciation?" IDI17, bilateral agency

One of the concerns raised by a respondent was that communicating HSS evidence was also challenging given its narrative complexity, and that stakeholders may consider adopting more creative approaches, such as communication grounded in personal narratives ("storytelling"), to reaching their constituents.

"I mean, I think our parliamentarians, the public don't understand systems. You know, they understand dying babies. People need to see something tangible. And so, there isn't a demand for it actually. Because then what there is…is we need to tell our story simpler. So, we often give people case studies. So we say, you know, it's the same health worker who's immunizing for COVID, as it is immunizing a child for measles..." IDI16, bilateral donors

3.4 How is evidence used in the decision-making process ?

A mix of factors were identified as involved in decision-making such as values, principles, donor priorities, and evidence. This relative mix varied considerably in the decision-making processes of specific global health funders. Discussing HSS in particular, some respondents took the view that more and better evidence was needed to justify requests to incorporate HSS into existing disease-oriented programs. Respondents differed in the motivation behind these evidence demands.

"With health systems, it's very hard to have a direct impact. And so, that's the problem. And then I think, if they hear pressure to do more on health systems in more cross cutting way, evidence, then or the lack thereof, gets used as the crutch for not doing it. I don't know that if there was robust evidence of the effectiveness of health systems, that it would really change. But I think the lack of robust evidence is a convenient excuse." IDI5, bilateral agency

…

" I still think for the Fund because they [donors] can say we've contributed, and we can tell you all about the lives we've saved. And it's a harder job when you start to talk about systems to be able to do that. And that's why we are so fixated a little bit on how can you show measurement of impact on health outcomes through systems investments, because fundamentally, that's what sells to a political realm around donor investment." IDI19, philanthropic organization

3.5 What are the perceived gaps in evaluating and tracking HSS investments ?

Need for methods to track impact of HSS investments: A few respondents raised the concern that the type of evidence sought by some global health investors around impact of HSS investments on disease outcomes was unattainable.

"People use words like health system strengthening… is that one thing? I don't think that's one thing. That's actually refers to a set of practices, approaches that is so heterogenous and diverse, it spans such a wide range of areas and approaches. And, you know, I think some people are on what's more than just a somewhat elusive quest. Yeah, evidence on what works to strengthen health systems? Like, I don't think if someone asked that question, you know, someone's sitting in like, whatever, like, you know, London or DC or Seattle is saying, where's my evidence on what works in strengthening health systems? I don't think that's an answerable question. Because, like, what part of the system are you trying to strengthen? And in what way and what context?" IDI3, multilateral agency

Assessing attribution or contribution of investments in HSS evaluations:

The respondents in this study shared divergent perspectives on the need for attribution or contribution in understanding the impact of their assessments on outcomes. Some donors felt strongly that such assessments were needed – for aforementioned return on investment concerns, as well as financial accountability reasons. Others felt that seeking attribution or contribution in the HSS space was methodologically challenging, due to the multi-dimensional nature of HSS investments.

"some of the measures that they want, we can't give them it would go against kind of the whole approach to health system strengthening, which is contribution." IDI2, multilateral agency

"I don't think it's possible, one. And, and I don't think it's important. Like I don't think you can sort of say, well, this much was [global health donor], this much was the [global health donor] But at the same time, if results are achieved, I mean, [global health donor] is going to say, but for us, those results wouldn't have been achieved. And I think it's fine if the country themselves say that, and the [another global health donor] says that right. So everybody can take full credit (…) I mean, one could argue that it is the full constellation of effort that achieves those results, but trying to separate out who owns what I mean, what's the point of it at the end of the day, right?" IDI5

"You know, and we need the attribution because of the weakness of those government PFM system. That's the ultimate…we're managing risk." IDI16, bilateral donor

Tracking financial flows: As indicated above, evidence of the impact of HSS investments was seen by a few respondents as important not only due to the need to understand impact on diseases, but also in terms of understanding corruption and misuse of finances. Previous experiences with HSS investments that resulted in misuse of funds might have also reduced the appetite for "blanket" HSS investment (such as in the context of SWAps) according to a few respondents.

"You need to know where to spend money, where you're doing it to protect and make sure that our money's not going somewhere where it shouldn't go. We really don't talk enough…of fiduciary risk, reputational risk, a corruption, okay, all that sort of stuff." IDI16, bilateral agency

"Because we don't believe you have the financial and programmatic assurance to know that the money is being spent in the right way." IDI19, multilateral

"[Country] has always committed to trying to build up a national ownership and leadership in its aid, and has, has always said that, at least. But then, obviously pays heed to the criticisms of the HSS investment era, or the, or the basket funding era. We didn't get anything out of it. It financed corruption or inefficiency, it didn't finance, couldn't finance determined action." IDI17, bilateral agency

"The sense was that [global health donors] didn't have that capacity to make sure this money was used in a smart way. But also, then to trace whether it had you know, where the money was coming in how it's being used, that it would just be used to top up salaries and go into this black box of which was some ways related to the at that time, the unease with these units, that was also the time of the SWAps. And it was coming towards the end of the SWAps, and the evaluations of the SWAps were not particularly positive in many

settings. So, these things have also combined to reduce this sort of collaboration between the agencies on our systems." IDI6, multilateral agency

Alignment in available HSS indicators: Looking beyond the impact of HSS investments on health outcomes, some respondents agreed that there was a body of work around indicators for HSS. However, challenges remain in applying them consistently across different contexts.

"People will talk about this lack of evidence and health systems or lack of indicators, and I'm always like, That's not true. There's a gazillion indicators. And the problem is there, none of them are comparable across countries or, you know, in the same country over time, because they're all very specific to whatever program was working on" IDI5

Another respondent noted that indicators were often designed or selected to meet the demands of funding cycles – and that certain aspects of the health system – such as the quality of health worker training programs – were more challenging to measure.

"So, there is a bit of a game played where you try and make your indicators as easy to obtain as possible...so we have a very sort of crude way of evaluating projects. And it is very simplistic, and it is open to gaming. So, you know, so this does temper the health systems piece of the work because things that are measurable, the program, things such as, you know, coverage and access things, which are more health systems such as improving institutional capacity, better curriculum, or a different curriculum for health isn't these are very much more binary." IDI6, multilateral agency

"We have more evidence of problems created, I would say, but less of results. And part of this may be a little bit of the time dimension of the grants. And I think that creates some bias. So Global Fund working on a three-year cycle. And you want to do system strengthening around like information systems and capacities to analyze and so on, it's not going to be a three-year program. So, I think that's really I mean, if they could, even if the grant cycle will stay the same, but they could embed that in a longer-term program of work that allows…I mean, even arguably, a five-year Bank cycle isn't really enough." IDI7, multilateral agency

A few respondents noted that the definitional challenges with HSS resulted in problems within organizations in terms of **internally** tracking and evaluating HSS investments.

"…it's just a bit of a nightmare when you start actually looking at the data and trying to calculate what is what can you classify as HSS versus something else, and everything overlaps…So once you start actually thinking about how to do it, just yeah, not very easy." IDI15, bilateral agency

"…we have [an evaluation unit], and they…looked at how, you know, health system strengthening can be evaluated, and they found that it's quite difficult, because it's not really defined, you know, what belongs to it." IDI12, bilateral agency

Aligning with the aforementioned preference for evidence by some in the donor community that showed an impact on outcomes, respondents noted that there has been a focus on methods such as RCTs.

"….still institutionally, it's much sexier to give us support the RCTs, get publications out saying it is rigorous evaluation work." IDI6, multilateral stakeholder

Areas for strengthening HSS evaluations Respondents described several areas for strengthening in HSS evidence and evaluations. A few respondents noted the need for more implementation research, locally contextual research, case studies, etc. that would allow for a deeper understanding of systems-level challenges, which would eventually lead to improved uptake of HSS investments.

"…there is a general bias towards positivist and quantitative methodologies, which are, of course, extremely useful and relevant in many instances. But then, you know, the policy analysis and case studies and small-scale implementation science is also equally important. And that is, of course, neglected." IDI1, multilateral agency

One of the issues with conducting context-specific implementation or operations research has been a lack of investment in building capacity for local implementing units.

"More recent attempts to say we will support local implementation research. And that's been put aside in projects as funding, but the uptake of that funding is very low, because of the capacity to country level to do it. And so there isn't there isn't a unit which is helping countries to put two hours together to put [proposals] together to help build capacity to do the other types of research, mixed method research, which there was for the randomized control trials." IDI6, multilateral agency

3.6 What are the gaps in the uptake of HSS evaluations?

Some respondents also discussed the challenges with utilizing existing evidence in the context of high workloads for staff members of global health organizations, such as length of reports, challenges with interpretation of results, and lack of evidence synthesis.

"I would love to be [someone] trying to digest a 100-page report, but I don't unless it's really important….we do these big evaluations over multiple months with all the data. And so you know, that that's hard to synthesize into a two pager…that's probably appropriate for the really, really, really big things. But how do we also share…snapshots and the key learning in real time as it's coming out? And…can we do ongoing evaluation that saw with a feedback loop that's feeding into program implementation, for implementation being adjusted, the evaluation being adjusted?" IDI14, multilateral agency

Finally, a few respondents believed that more translation and dissemination of evaluation evidence – including HSS evaluation – is needed at the country-level, and that evaluations currently appear to be tailored to donor uptake, as opposed to country-level stakeholders. Local researcher groups within LMICs were also not sufficiently engaged in the research process, according to one respondent.

"…we don't see a lot of translation of all [this] knowledge…we see rarely, the evaluators coming back to the countries and give back the results…" IDI13, national-level stakeholder

"… the message broadly given to country says, give us what we are interested in rather than what you're interested in." IDI18, researcher

3.7 What are the barriers to improved coordination around HSS evaluations ?

Misaligned incentives of global health investors: Respondents broadly held the view that the incentives for donors to cooperate on HSS investments and evidence-generation could be better aligned. One of the underlying drivers for the lack of alignment across many global health priorities was the different structures of accountability of donors to their own constituencies.

"…each institution, the internal incentives and disincentives are extremely strong and quite distinct from each other. So, the alignment of incentives across organizations is extremely difficult to get." IDI6, multilateral agency

"*And it's ironic because they're the same investors across several of us health priorities. And so, I actually feel a lot of this if they could sort it out themselves what they wanted things, they could very easily through our governance mechanisms cascade that down to us, but they don't. And what they do is they pressure us to try to come up with some of these answers that us in silo can't come up with." IDI2, multilateral organization*

Lack of coordination within organizations: A few respondents noted that generating internal consensus around definitions and scope of HSS within organizations would be an important step in improving overall coordination of HSS evaluation efforts.

"Because I think if you just engage the health system strengthening folks at [global health donor], you're going to miss a lot of what happens, and also the potential for what happens. Should they be better aligned? Yes. Are they? I don't think so." IDI20

"…getting donors to reach consensus on this. So I've been trying for five years, on HSS. And it's almost impossible. And I haven't given up. And you need to give them some hooks, we need to get people hooked on it. But it's only the technical folks that come up with the ideas, and we then have to sell it to our ministers. And if you've got a minister who doesn't really know about HSS, and doesn't really care about HSS, we've got to sell it there before it can then happen at that level. You know, there's the willingness at this level, but you know, they'll, again, they're responding to taxpayers. You know, where are the vaccines? are they what are happening with them? And, you know, this is this is where the traction and getting that's really difficult." IDI16, bilateral agency

Misalignment between donors and national governments: Some respondents also believed that donors and LMIC governments shared divergent views. The reasons for divergences differed amongst the respondents. One respondent noted that donors and country governments were not aligned due to the fact that countries saw their systems as integrated, while global health investors continued to favor a 'vertical' or 'disease-oriented' funding approach. Two other respondents believed that country priorities were in fact not homogenous; expressed by multiple 'internal' constituencies. In other words, perspectives on national priorities could differ substantially based on who donors were primarily engaged with at the country-level. These various national level units could also utilize evidence to underscore investments in their particular space.

I"I would very keen to see the bilaterals, the multilaterals and the global alliances, adopt a much more flexible approach, because this is generally at variance with what countries want. I mean, countries look at their health systems as very integrated health systems that

are meant to provide care for all who need it, rather than the few who are program recipients ." IDI18, researcher

"if you're talking to a Minister of Finance, he wants the economic case. And he's got the purse. So that's the evidence he wants. Ministry of Health tends to, you know, be in different directorates. So they wanted to protect the investments going into their directorates. So I think it really depends on the type of evidence, you know, I think I think countries are really savvy in terms of, you know, using and giving and asking for different types of evidence when they need it." IDI16, bilateral donor

"Most of the decisioning most of the financing decisioning is still happening in western capitals, amongst donors, and according to donors' interests, and, and those interests are very, very valid. But if you want, so if you interpret health system strengthening work in the political sense of reorienting aid, to no longer go in and do and say how many women are getting C sections in reasonably well functioning EmOC centers, or how many kids are vaccinated with DPT-3 or how many women's unmet need for FP is, you know, if you want to say, Okay, we're going to come in and we're going to support this country, to develop their healthy ecosystem, as much as possible to be both equitable and affordable and with a reasonable degree of quality, we're going to let the sovereign state drive, drive the decision, then, then we're very far away from it." IDI17, bilateral donor

Related to the issue was a statement from one respondent who indicated that increased focus on HSS without the accompanying requests for impact evaluation – could be achieved with greater country ownership.

"But ultimately, if you're thinking about where one wants to get… to try to cut this attribution contribution argument off at the knees, is for the countries to be telling the donors what they need and what to do." IDI19, philanthropic organization

The limited space for LMICs in the issue of evaluation uptake appeared to reflect challenges in expanding the voice for these countries within global health initiatives.

"So frankly, this conversation tends to be dominated by donors and not implementers on the board." IDI20, philanthropic organization

3. Overarching Findings/Themes

Global health stakeholders rely on – and demand – different types of evidence in HSS evaluations, and for different purposes. Greater clarity in categorising these different types of evidence would be beneficial.

Funding for global health is finite, and investors need to balance multiple dimensions in their decision-making: impact of investment on health outcomes, time to impact (lives saved in the near-term v. the long-term), perceptions and priorities around the disease- or health-conditions targeted, and national interests on the part of bilateral donors.

Global health investors diverged considerably in the type of HSS evaluation evidence sought or desired, and in the end purpose for gathering that evidence. Some respondents sought evidence that draws a "line of sight' between HSS investments, outputs and outcomes, including impact on health status. Other investors want to understand the overall cost-effectiveness of HSS investments compared with disease-specific investments, as well as relative "cost-effectiveness" of HSS interventions (investments in health workforce compared with investments in supply chain systems). Finally, a few respondents stressed the need for context-specific evidence that examined major learnings and considered how to improve HSS investments (understanding the "how").

Despite a number of global agreements that evaluating the **contribution** of investments to collective outcomes rather than **attributing** specific outcomes to specific investments is desirable, respondents noted that funders were still often interested in attribution.

The uptake of HSS evaluation evidence was complicated by problems with timeliness and interpretability. A few respondents shared perceived challenges with interpreting evidence on HSS, likening some health systems work as akin to "art appreciation". Others noted that HSS evaluation timelines did not match programming cycles, and said that more emphasis on real-time evidence would be useful to investors. Communicating HSS evidence was widely acknowledged as challenging given its complexity. In addition, some respondents believed that more translation and dissemination of HSS evaluation evidence was needed at the country-level, and that evaluations currently appear to be tailored to donor uptake, as opposed to national-level stakeholders. It was clear that the HSR community can do better in communicating its findings in ways that are clear and actionable (for policy makers) and are also clear and understandable on a more human level (how do health systems impact on people's lives?). Weaknesses in consistent policies towards HSS by donor agencies, poor definitions and a lack of funding have all contributed to weaknesses in the development of this field.

Barriers to a coordinated approach to HSS evaluations included misaligned values, incentives and architectures among global health donors, within donor organizations and between donors and national-level stakeholders: . Internal consensus within global health donors regarding health systems and HSS also stood as a barrier to coordinating evaluation efforts. Finally, evaluation questions were often driven by donors and not national-level stakeholders, creating further barriers to effective coordination across all stakeholders.

4. Annexes

Methodology

This study was designed as a prospective stakeholder analysis, utilizing qualitative methods. Qualitative methods such as interviews are commonly utilized in stakeholder analyses in order to understand individual and organizational perspectives on particular issues, and to delve into topics such as power, interests and relationships. Stakeholders in this study were defined as a) **investors in global health programs**; b) **national-level decision makers or implementers**; c**) researchers, civil society or academics** involved in HSS investment processes at the global-level. This study was conducted by Itad staff (Natasha Palmer, Shreya Pereira) and external research consultants (Veena Sriram, Sara Bennett, Krista Kruja).

Box 1: Examples of interview questions

How would you describe how global health funders view the types of evidence needed to assess the value of their HSS investments?

In your opinion, what are the current gaps in how global health organizations conceptualize and commission HSS evaluations?

What is the type of evidence that you think global health funders *want* to see that would facilitate an increase in HSS investments?

What can be done to improve the engagement between researchers and

Data collection

This study drew upon semi-structured interviews as the primary source of information. We supplemented interviews with document review from other workstreams.

Sampling: We utilized purposive sampling to select our respondents (Patton 1990). The first step was to develop a master list of potential respondents from the following categories : 1) multi-lateral agencies ; 2) bi-lateral agencies ; c) philanthropic organizations ; d) civil society organizations ; d) national health authorities ; e) research organizations. The master list emerged through extensive discussions amongst the research team and suggestions of the HSSEC executive committee and working groups. We also drew on suggestions from respondents involved in the study. Sampling decisions were taken to ensure diversity across types of stakeholders, types of global health investors and positions within the organization.

Instrument Guide Development: Guides were developed for the following groups : 1) global health investors ; 2) research groups and civil society ; 3) implementing country stakeholders. Guides were pilot tested with two respondents and then periodically revised to reflect learnings from the interviews. Given the semi-structured nature of the interview process, flexibility was given to interviewers to raise probes or follow-up questions building from responses given within the interviews. Box 1 presents sample questions.

Implementation: Interviews were conducted in pairs by Veena Sriram, Natasha Palmer, Shreya Pereira and Sara Bennett, using the Microsoft Teams or Zoom platforms. Verbal consent and permission to record was sought from all participants. Verbal consent documents and categories of questions were provided to the respondents in advance of the interview.

We conducted 23 interviews with 25 respondents (two interviews were conducted with jointly two respondents), as described in Table 1. Further information regarding organizations or organizational affiliations are not provided, in order to protect respondent identities.

Table 1: Stakeholder categories and interview respondents

Stakeholder category	Number of interview respondents
Multilateral agencies	8
Bilateral agencies	7
Philanthropic organizations	6
Civil society organizations	1
National-level stakeholders	1
Research organizations	2
Total	25

Data analysis

Audio recordings from interviews were transcribed using otter.ai, and then cleaned and checked by external consultants (Krista Kruja, Mel Michener). Initial themes were developed by the research team through multiple reviews of transcripts (Veena Sriram, Krista Kruja, Mel Michener) and regular debriefing within the team. To further develop specific themes, a framework approach to qualitative data analysis was utilized to facilitate deeper analysis (Gale et al. 2013). Interim findings were shared amongst the research team, and with members of the Executive Committee, and feedback was utilized to strengthen and modify existing themes.