

The Devil Is in the DLIs: What Program-for-Results Health Operations Can Teach the State Department About Results-Based Financing

 Jocilyn Estes and Rosie Eldridge

Recent [CGD case studies](#) draw lessons from three performance-based, on-budget health assistance projects—including the World Bank’s first Program-for-Results (PforR) project in Ethiopia—to inform the State Department’s new bilateral health agreements. That analysis yielded several insights into the design of performance-based programming, adding to years of existing CGD research on [paying for what works](#) in foreign assistance.

Now, as State Department officials craft implementation plans for more than 30 bilateral agreements signed with partner countries, the question of how to design results-based indicators is more urgent than ever. We think the devil is in the DLIs—the disbursement-linked indicators, that is. The America First Global Health Strategy aims to pay for results, but designing such instruments can be onerous and time-consuming, and if poorly designed, implemented, or measured, this new approach could just mean new disruptions to care for [patients around the world](#).

The World Bank’s PforR instrument offers over a decade of experience to learn from. We built a [dataset](#) covering the World Bank’s PforR health portfolio to assess what these operations actually measure, how they perform, and what lessons they hold for this new iteration of results-based health financing. The State Department should adopt a concise set of clearly defined indicators, balancing achievable milestones with more ambitious health outcomes through a portfolio-style approach that builds on existing government systems. Strong verification processes, flexible approaches to challenging reforms, and transparent performance tracking from the start will be essential to translating funding into sustained health gains.

Background on PforR instrument

PforR is the newest of the World Bank's three financing instruments. The first commitments were made in FY2012, and its use has only grown since then. It provides program support through countries' own systems and disburses funds only when defined results—captured through disbursement-linked indicators (DLIs)—are achieved and verified.

A typical PforR project defines indicators and targets negotiated between the World Bank and the partner government before implementation begins. During implementation, an independent entity verifies whether results have been achieved before releasing payment. The World Bank documents these operations through appraisal, progress, and completion reports—referred to as PADs, ISRs, and ICRs—which formed the basis of our data collection.

Methodology

To assess what PforR health projects measure and how they have performed to date, we built a [dataset](#) of all PforR operations with at least 50 percent health sector focus, based on the World Bank's sector classification. The sample comprises 46 PforR projects—16 closed, 30 active—approved between 2013 and 2026, with 475 individual DLIs. We excluded 15 follow-on funding operations to avoid double-counting DLIs.

We reviewed documentation across all 46 projects, manually recording each DLI and its achievement and disbursement status. Achievement was coded as achieved, partially achieved, or not achieved for closed projects, with off-track and not assessed added for active projects. Disbursement was coded as full, partial, or none, with not assessed added for active projects. The disbursement analysis focuses largely on closed projects.

We then classified each DLI in two ways:

First, we assigned each DLI one of the six [WHO health system building blocks](#): service delivery, health workforce, health information systems, access to essential medicines, financing, and leadership/governance. Classification was based on what the DLI directly measures, not what it ultimately aims to improve.

Second, we adopted the taxonomy used in earlier CGD analysis—Gelb et al.'s [Program for Results: The First 35 Operations](#) (2016)—to place each DLI on a spectrum from process to results based on what the borrower must do to trigger disbursement (Table 1). Gelb et al. applied this taxonomy across all sectors, but here we apply it to health specifically.

For the seven projects overlapping with the earlier CGD analysis, we adopted the published classifications. For all other projects, we used Claude Opus 4.6 (Anthropic) to generate initial DLI

Type and Health Domain classifications based on DLI descriptions, the Gelb taxonomy, and the WHO building blocks framework, and to verify DLI Type classifications against PAD verification protocols. Gelb et al. note that “an element of judgment is often necessary” in classification, which we addressed by spot-checking across all categories and revising 39 of 475 classifications through systematic review.

Table 1. Disbursement-linked indicator types

DLI TYPE	ABBREVIATION	DEFINITION	EXAMPLE
Input	I	Evidence of procurement or spending	(None in health portfolio)
Action	A	A specific, discrete policy action	Adoption of revised national health strategy
System Action	SA	Complex, multi-step institutional strengthening	Design and implementation of performance-based incentive scheme
System Output	SO	System performance or capability measure, upstream from service delivery	Percentage of facilities meeting quality standards
Output	O	Delivery of a specific service or product to beneficiaries	Children immunized with Penta 3 vaccine
Outcome	OO	Longer-term achievement requiring behavioral change beyond service delivery	Hypertension patients with blood pressure controlled

Health PforR by the numbers

The health PforR portfolio comprises 46 projects—16 closed and 30 active—totaling \$13.6 billion in financing across 475 individual DLIs. Projects range from \$9.3 million (Samoa) to \$1.2 billion (Indonesia COVID), averaging 10 DLIs and roughly \$31 million per DLI.

Table 2. Summary statistics for the health PforR portfolio (2013–2026)

METRIC	
Total projects	46
Closed projects	16
Active projects	30
Total DLIs	475
Health-relevant DLIs	459
Non-health-relevant DLIs	16
Fewest DLIs	Colombia (4)
Most DLIs	Pakistan (27)
Total PforR financing	\$13.6 billion
Mean project size	\$296 million
IBRD commitment share	66%
IDA commitment share	29%
Smallest project	Samoa (\$9.3 million)
Largest project	Indonesia COVID (\$1.2 billion)

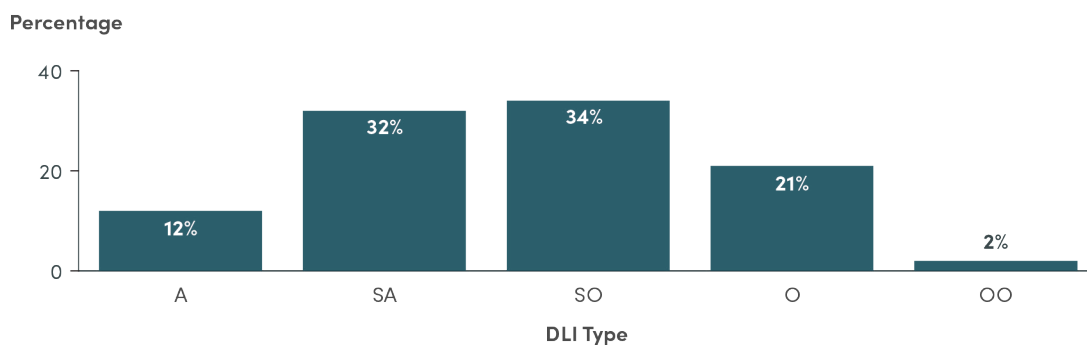
PforR health operations largely pay for process, not results

Classifying each DLI on the Gelb et al. spectrum reveals a portfolio of indicators that largely rewards process accomplishments, not actual results. Discrete policy actions (A) account for 12 percent of all DLIs. Complex institutional reforms (SA) and system performance measures (SO) together account for 65 percent. Only 21 percent are outputs (O) measuring services delivered directly to patients, and less than 2 percent—just 8 DLIs across the entire portfolio—measure health outcomes (OO) that require behavioral change.

Figure 1. What are health PforR operations actually paying for?

DLI Type distribution across 475 DLIs in 46 World Bank health PforR operations

Indicators are classified as Input (I), Action (A), System Action (SA), System Output (SO), Output (O), and Outcome (OO). There were no indicators classified as inputs in the health portfolio.

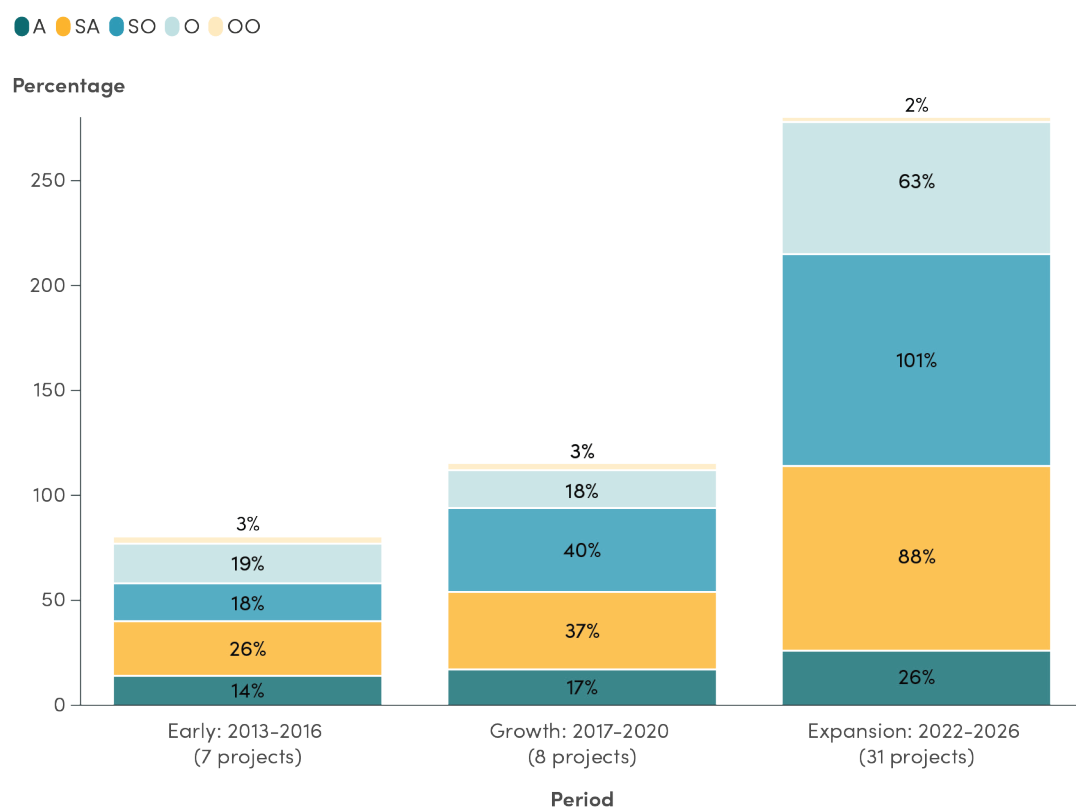


This pattern has deepened as the portfolio has grown. The portfolio has tripled in size—from 80 DLIs across 7 projects in 2013–2016 to 280 DLIs across 31 projects in 2022–2026. Over this period, the share of discrete policy actions (A) triggering disbursement declined by half, from 18 to 9 percent, while system performance measures (SO) rose from 23 to 36 percent to become the single most common DLI type. Complex institutional reforms (SA) held steady at around a third, and outputs (O) at roughly a quarter. Most striking is the consistent scarcity of DLIs tied to health outcomes (OO), whose share fell from 4 to just 1 percent.

Overall, PforR health operations have expanded, concentrating on system performance (SO) while rarely linking funding to health outcome (OO) measurement.

Figure 2. Has PforR design evolved over time?

Number of DLIs by type and approval period, 46 health PforR operations



These portfolio-level numbers for DLI type mask variation across projects. Some focus on patient-level results; eight projects contain no patient-level output or outcome indicators at all. This variation suggests that the PforR instrument can accommodate indicators at a variety of points on the results chain, the choice of indicators is ultimately a design decision shaped by other incentives and constraints.

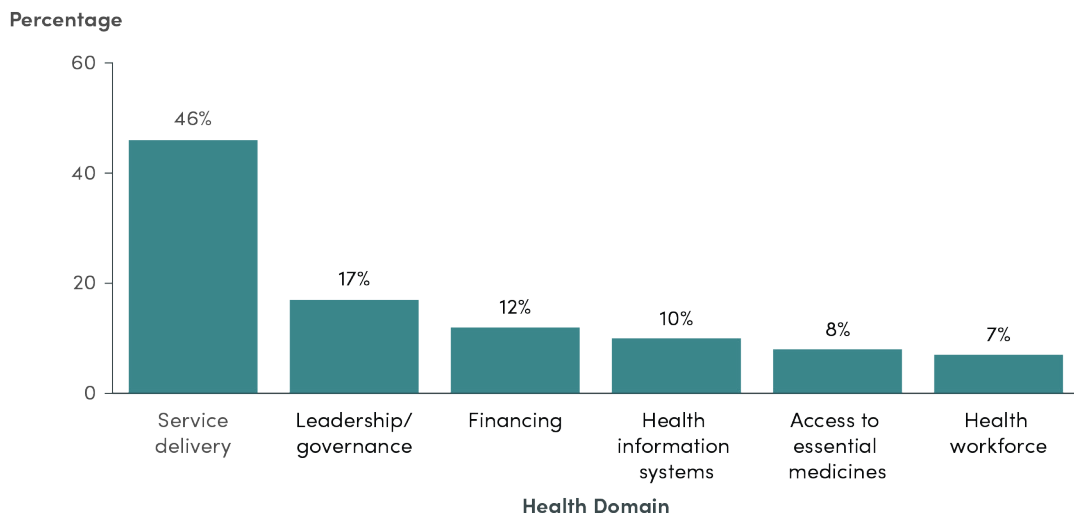
Why health PforR operations concentrate on process indicators—when the instrument was designed to pay for results—is a question this analysis raises but cannot fully answer. The World Bank’s own [Independent Evaluation Group offers one explanation](#): DLI design tends to favor indicators that disburse predictably over ones that stretch performance. Process indicators are easier to define, measure, and verify, making them a more practical option.

Nearly half of all DLIs focus on service delivery

By health domain, service delivery accounts for nearly half of all health DLIs (46 percent), with the rest spread across governance (17 percent), financing (12 percent), health information systems (10 percent), essential medicines (8 percent), and health workforce (7 percent).

Figure 3. Where in the health system are PforR resources concentrated?

WHO Health Domain distribution across 459 health-relevant DLIs in 46 World Bank PforR operations



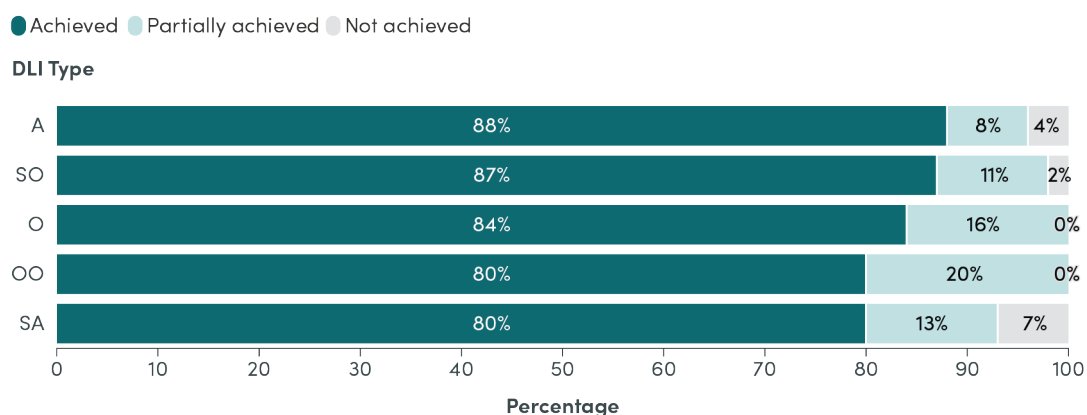
Simpler indicators are more reliably achieved

Across the 16 closed health projects, 84 percent of DLIs were fully achieved. However, this figure masks important variation in DLI type and health domain.

Achievement rates vary modestly by DLI type, but a pattern holds: the simpler the indicator, the higher the achievement rate. Discrete policy actions (A) are the most reliably achieved at 88 percent. System performance measures (SO) and outputs (O) are close behind at 87 and 84 percent. Complex institutional reforms (SA) and health outcomes (OO)—both requiring sustained effort over years—sit at the bottom at 80 percent each, with SA carrying the highest outright failure rate (7 percent).

Figure 4. Does DLI type affect achievement?

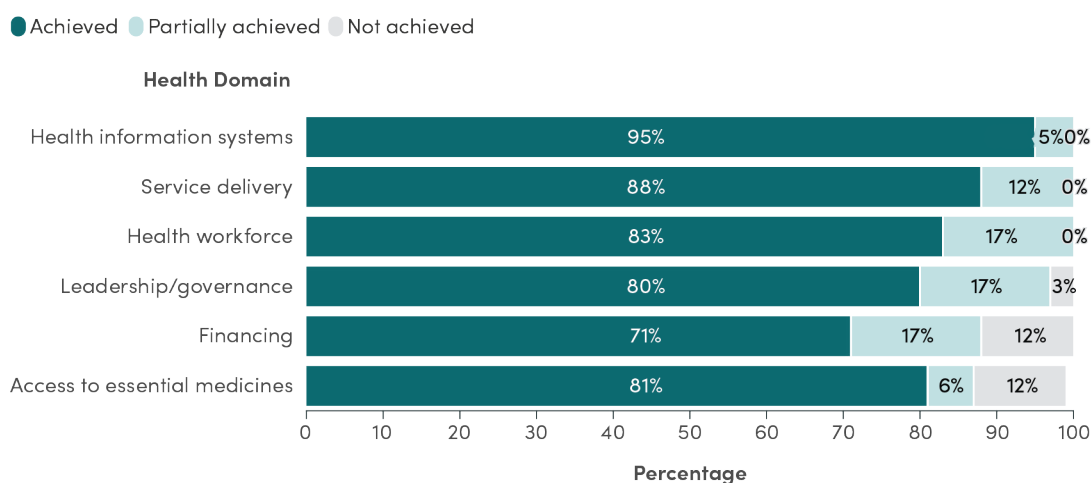
Achievement rates by DLI Type for 16 closed health PforR operations (170 DLIs)



A different pattern emerges across health domains. Health information systems stand out at 95 percent, followed by service delivery at 88 percent. But financing reforms achieve at only 71 percent—the lowest of any domain—and access to essential medicines has the highest failure rate (12 percent).

Figure 5. Does health domain affect achievement?

Achievement rates by Health Domain for 16 closed health PforR operations

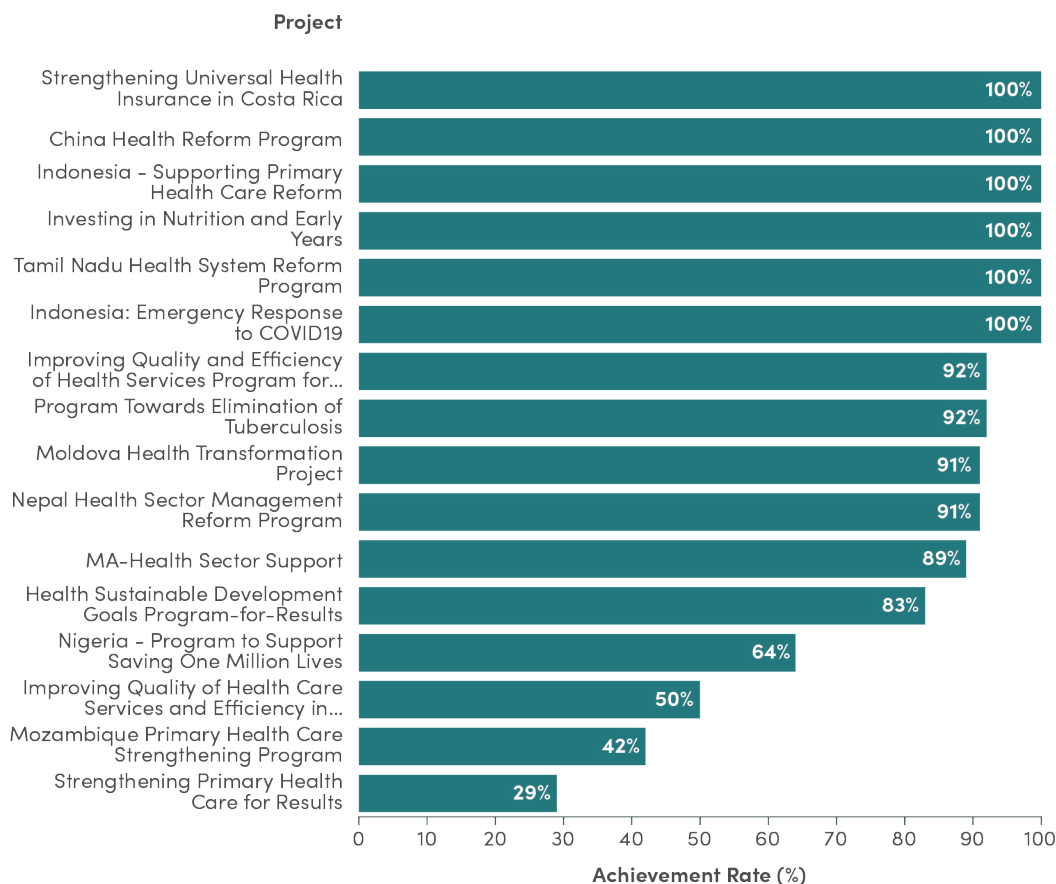


Project-level achievement varies widely

Project-level achievement varies widely, from 100 percent to 29 percent. Six of 16 closed projects achieved every DLI, while three achieved less than half.

Figure 6. Which projects achieved their DLIs?

Percentage of DLIs fully achieved per project, 16 closed health PforR operations

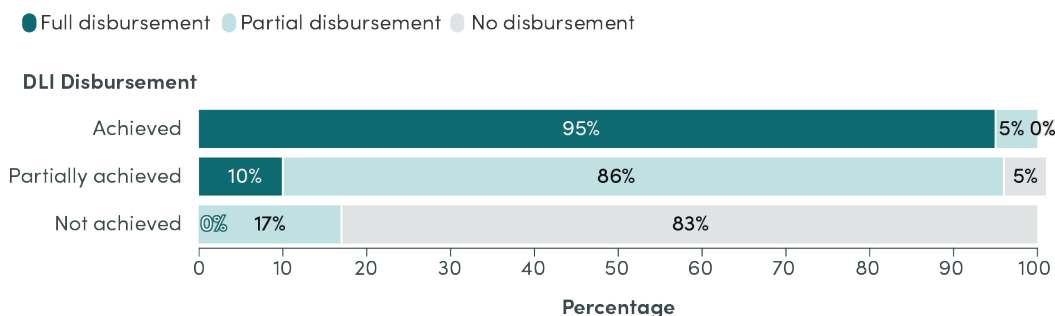


The disbursement mechanism largely works as designed

Ninety-five percent of achieved DLIs received full disbursement, 86 percent of partially achieved DLIs received partial payment, and 83 percent of not achieved DLIs received no disbursement. The few mismatches may reflect verification timing, sliding scale provisions, or financing reallocations during restructuring. That said, “partially achieved” covers an enormous range, from near misses to significant shortfalls, and the disbursement implications vary just as widely.

Figure 7. Does achievement lead to disbursement?

Disbursement rates by achievement status for 16 closed health PforR operations (170 DLIs)

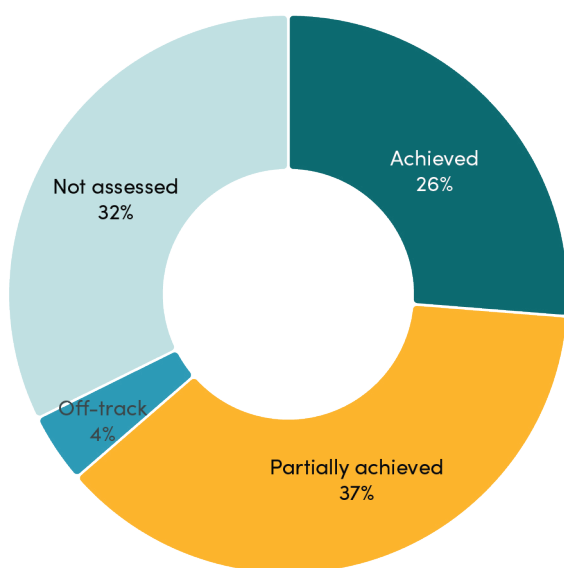


Most active projects are showing early progress

While much of what we can deduce about PforR health projects is based on closed projects with final results, it is important to track the trends emerging from projects currently underway. Across the 30 active projects, 63 percent of 305 DLIs show progress, with 26 percent fully achieved and 37 percent partially achieved. Only 4 percent of DLIs (13 total) are explicitly off-track. However, nearly a third of DLIs remain too early in implementation to evaluate. Because active projects make up the majority of the PforR health portfolio, the full picture of health PforR performance is still unfolding.

Figure 8. How are active projects progressing?

Achievement status of 305 DLIs across 30 active health PforR operations



Gelb et al.'s finding holds even more strongly in health

In 2016, Gelb et al. analyzed the first 35 PforR operations across all sectors and asked whether PforR was delivering on its promise of linking financing to results. Their analysis revealed that most projects did not extend very far down the results chain. A decade later, our analysis of 46 health operations tells a broadly similar story.

Table 3. PforR health portfolio relative to Gelb et al. analysis

METRIC	GELB ET AL. (ALL SECTORS, 2016)	THIS ANALYSIS (HEALTH, 2026)
Portfolio statistics		
Total financing	\$8.1 billion	\$13.6 billion
Projects	35	46
Total DLIs	278	475
Average DLIs per project	8	10
Average \$ per DLI	\$30 million	\$31 million
DLI type distribution		
Actions (A + SA)	54%	43.8%
System performance (SO)	17%	33.5%
Results (O + OO)	29%	22.7%

Compared with Gelb et al.'s all-sector portfolio, health PforR operations have shifted away from action indicators, but this gap has been filled by system performance measures rather than patient-level results. System performance (SO) indicators account for 34 percent of health DLIs—double the 17 percent share in Gelb et al.'s all-sector portfolio—while results-oriented DLIs (O + OO) are actually lower in health at 23 percent than the all-sector average of 29 percent. Most strikingly, health outcome DLIs have become rarer, not more common, with only 8 of 46 health projects to include even one health outcome (OO) indicator.

Lessons for the State Department

The World Bank's PforR instrument and the State Department's bilateral health agreements differ in important ways. The World Bank has extensive experience working with governments and managing financial risk; lending creates different incentives and leverage than US grant assistance; and the PforR instrument isn't built around health system graduation timelines. But both focus on results-based financing, and more than a decade of PforR health operations offers several lessons.

Designing the indicators

- **Keep the indicator list short and focused.** The bilateral health agreements span a wide range of program areas, which could create pressure toward long indicator lists. The PforR health portfolio we assessed averages 10 DLIs per project, roughly \$31 million per indicator, with a wide range from 4 to 27 DLIs. More indicators mean more verification requirements that are often costly and only loosely linked to outcomes.
- **Separate what triggers payment from what is measured.** Not every indicator needs to be tied to payment. In PforR operations, only a subset of indicators (DLIs) trigger disbursement, and the rest are tracked for program management. The State Department should take the same approach. But tracking without consequences is not enough. A World Bank evaluation found that indicators not tied to payment were frequently neglected. The State Department should build accountability and allocate sufficient resources around its non-payment indicators—and accompanying data systems—to avoid the same pattern.
- **Be explicit about what each indicator measures.** Many PforR DLIs use broad or ambiguous language, with the specific result that triggers payment buried in verification protocols. Each bilateral agreement indicator should clearly and transparently state where it sits on the results chain, what triggers payment, and how it connects to the partner government's own health strategy.
- **Think carefully about what each indicator is worth—and think creatively about funding the most important milestones.** The World Bank does not tie indicator values to the cost of achieving them. A DLI worth \$30 million does not mean it costs \$30 million to achieve. Instead, valuation is strategic, with higher-importance or foundational indicators signaling priorities and creating stronger incentives. In contrast, the State Department's assistance instruments do not generally allow for the flexibility to pay much beyond a costed baseline for milestones achieved. Despite this crucial difference in what the US government can pay for, the World Bank's experience suggests that for key indicators or milestones with high strategic value, the State Department should consider additional technical assistance or other support to supplement performance-based instruments. The State Department must also be deliberate in how it distributes financing across indicators—including which to link to payment at all—even while limited by available procurement modalities.

- **Process indicators can drive health gains, if designed intentionally.** Health outcomes (i.e., health gains achieved) are the goal, but they are difficult and expensive to measure accurately, take longer to materialize than outputs (i.e., health services delivered), and often reflect multiple contributing factors. But process and output indicators can still drive health gains without a direct, easily verifiable link to patient outcomes—particularly when those DLIs are focused on improvements to existing actions or processes. In the Ethiopia PforR case study, systems-strengthening indicators aligned with existing local frameworks were judged to drive health gains. Given its limited staff and verification capacity, the State Department should link payments to a small number of easily authenticated, potentially catalytic systems-level indicators.
- **Don't let ambition disappear from indicator design.** Health outcome indicators have dropped to just 1 percent of the PforR health portfolio—only 8 of 46 projects include even one such metric. While process indicators are practical given current constraints, the State Department should not abandon the other end of the results chain where feasible. The State Department should aim to include at least some indicators tied to patient health outcomes where aligned with a government's existing national health plan. These DLIs should be measured over a longer time horizon to allow sufficient time for impact to materialize and be reliably assessed.
- **Finally, build on existing indicators rather than reinventing the wheel.** Many PforR operations already pool funding from various donors—including trust funds—through a single set of DLIs, aligning multiple financing streams around one cohesive set of results. Where the State Department's bilateral agreements overlap with existing PforR operations, it should use existing indicators and verification arrangements rather than creating parallel frameworks. Co-financing existing indicators would reduce fragmentation, limit the reporting burden on partner governments and US staff, and avoid duplicating work that has already been done.

Managing performance

- **Balance achievable wins with ambitious goals.** Achievement rates vary by indicator type and health domain, so the mix matters. Weighting the portfolio toward indicators with relatively high achievement rates and low verification costs allows money to flow, while freeing up government and State Department capacity for the more difficult indicators that are critical to strategic reform goals. Payment-linked indicators should be largely output-based and quantitative, with targets set based on baseline trend data and a sliding scale that rewards partial credit, rather than binary thresholds that risk withholding payment entirely.
- **Handle co-financing and commodity handovers with care.** The bilateral health agreements focus heavily on domestic co-financing and the transition of essential commodity procurement—two areas that have drawn significant criticism and where PforR indicators struggle most. In our analysis, financing reform DLIs had the lowest achievement rate at 71 percent, and access to

essential medicines indicators had the highest failure rate at 12 percent. Rather than linking rigid payment deadlines to these areas, the State Department should treat co-financing as trajectory milestones with built-in flexibility and carefully sequence essential commodity handovers to avoid disruptions to care. The Global Fund's Sustainability, Transition, and Co-Financing Policy offers one example: it unlocks 15 percent of a country's allocation when co-financing commitments are met, with flexibility in how those commitments are fulfilled.

- **Design verification processes to mitigate risk.** In our analysis, multiple projects met their DLI targets but experienced delayed payment because verification was late or incomplete. Verification protocols should be designed before implementation begins, ideally drawing on existing government data systems. Implementation plans should include fund rollover allowances for verification delays, and verification itself should be explicitly budgeted. The State Department should also consider advance financing provisions—PforR operations can advance up to 25 percent of financing before results are verified, providing governments with the liquidity needed to implement before demonstrating results.

Tracking for transparency

- **Standardize reporting across agreements.** Across the PforR portfolio, every project reports indicators differently—with inconsistent detail, frequent gaps, and contradictions between data fields and narrative text. Even achievement labels lack consistency—"partially achieved" can mean anything from a near-miss to a significant shortfall. Restructuring further complicates the picture, as final reporting does not clearly distinguish between original and restructured indicators. The State Department should adopt a standard reporting template from the outset and report achievements as percentages of targets where relevant. When indicators are revised or dropped, new ones should be reported alongside original targets with a written explanation.
- **Pay for data quality.** Health information system DLIs had the highest achievement rate in the portfolio at 95 percent. Tying payment to better reporting and data audits improves the government's systems while strengthening the State Department's ability to verify every other indicator. Where tying payment to a data system milestone isn't appropriate, the State Department should still invest in data quality through traditional assistance mechanisms.
- **Build a performance database from day one.** No central database of PforR DLI performance exists. This dataset was assembled from hundreds of individual project documents. With more than 30 agreements signed but awaiting implementation, the State Department can do better from the start. A public performance database, regularly updated with indicator progress, verification status, and disbursement across agreements, would enable real-time learning, course correction, and public accountability.

The World Bank's decade-plus of PforR health operations shows that results-based financing can successfully disburse funds and drive reform, even as the vast majority of its indicators measure system-level processes rather than patient outcomes. Despite its extensive government partnerships, established public financial management tools, and technical capacity, less than 2 percent of the World Bank's health PforR indicators are tied to health outcomes. The State Department will face an even steeper challenge as unprecedented transition timelines risk outpacing the institutional changes needed to sustain results. With less implementation experience, fewer staff, and less leverage than the World Bank, the State Department should prioritize credible, verifiable milestones aligned with country systems—while not losing sight of health outcomes over time. Getting this balance right will require realistic timelines, adequate technical support, and investment in the data systems needed to measure performance. A decade of PforR health operations offers a roadmap for getting there.

With thanks to Janeen Madan Keller and Phillip Palmer for their helpful comments and suggestions.

JOCILYN ESTES is assistant director of US Development Policy and a senior policy analyst at the Center for Global Development.

ROSIE ELDRIDGE is a research associate at the Center for Global Development.

Suggested citation:

Jocilyn Estes and Rosie Eldridge. 2026. "The Devil Is in the DLIs: What Program-for-Results Health Operations Can Teach the State Department About Results-Based Financing." CGD Note 409. Washington, DC: Center for Global Development. <https://www.cgdev.org/publication/devil-dlis-what-pforr-health-operations-can-teach-state-department-about-results-based>



www.cgdev.org

This work is made available under the terms of the
Creative Commons Attribution-NonCommercial 4.0 International License.

