POLICY BRIEFS ON

ECONOMIC IMPACT OF HIV

16.

TRADE-OFFS WITHIN THE HIV BUDGET

Health Economics and Epidemiology Research Office
HE²RO
Wits Health Consortium
University of the Witwatersrand

WITS UNIVERSITY

BOSTON UNIVERSITY

UCL

This brief forms part of a body of work on the Economics of HIV, funded by the Bill & Melinda Gates Foundation (INV-002382). The authors acknowledge the contributions of the participants of the 'Economics of HIV' meeting in Cascais, Portugal, in September 2018 for general direction on this work, and thank participants of a webinar in March 2021 for comments on a previous version of this brief. Additionally, the brief was reviewed in-depth by Tim Hallett from Imperial College London. We are grateful for the excellent work of James Baer in proofreading the briefs, Carla Hauptfleisch in designing them, and Michael Obst in research support. The findings and conclusions contained within this brief are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation or of the institutions the authors represent.

Recommended citation:

Gesine Meyer-Rath, Kate L Harris, Markus Haacker: Trade-offs within the HIV Budget. Policy brief #16 of series "Economic Impact of HIV". Johannesburg, December 2021.

16.

TRADE-OFFS WITHIN THE HIV BUDGET

KEY POINTS

- **Governments need to choose between different HIV interventions in order to maximise impact under limited funding. This can be done by ordering interventions by cost-effectiveness in a league table, or by employing allocative efficiency models to determine how to allocate funds across packages of interventions in the most cost-effective way.**
- **There can be a tension between cost-effectiveness (often expressed as cost per health outcome achieved, such as HIV infections averted or life-years saved) and other optimisation targets such as equitable access to services, or ending AIDS as a public health threat. Trade-offs also apply over time, as increasing the coverage of interventions today improves health outcomes but also affects spending needs in the future.**
- **Decision-making is complicated by (a) declining returns to investment in individual interventions, (b) increasing marginal costs at high coverage levels, (c) interactions in effectiveness between different interventions and (d) a lack of effectiveness data for some interventions.**
- **Additional practical constraints to achieving allocative efficiency include political promises from local and global actors, and budgetary inertia that makes drastic shifts across years difficult. These and the need for large amounts of data (and the risk of injurious advice given the uncertainty in the data) limit the usefulness of optimisation models, particularly for HIV programmes at high levels of intervention coverage.**

Even within the HIV budget, choices are necessary to sustain or increase progress towards AIDS control – in this case, trade-offs between different interventions. In the context of a global HIV funding envelope that has stagnated for the last ten years, governments and donors emphasise allocative efficiency in HIV programming – in other words, the identification of the mix of HIV interventions that produces a defined level of output at the lowest possible cost or achieves maximum results within a given budget constraint.

Traditionally, the economic value of interventions is analysed considering one intervention at a time. In order to establish the allocative efficiency of an entire programme with many different potential interventions, models often integrate additional aspects such as the impact of one intervention on another (e.g., of prevention on treatment), non-linear effects of different coverage levels, spatial targeting and additional objectives such as equitable coverage or epidemic control.

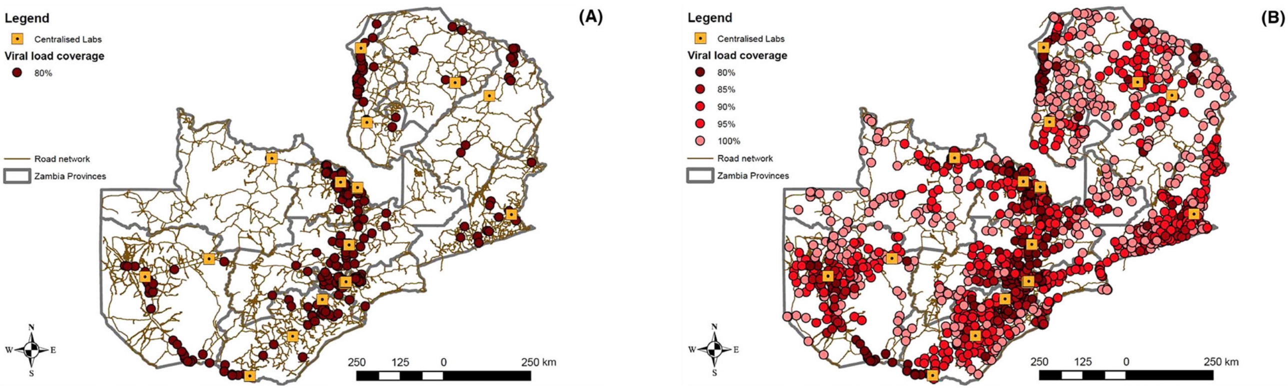
The role of allocative optimisation models

Governments need to choose between different HIV interventions. One traditional approach is to order interventions by cost-effectiveness and allocate budget to the most cost-effective interventions. This can be supported by models that optimise the allocative efficiency of the HIV budget.

Allocative efficiency optimisation models in HIV aim to optimise spending allocations to the HIV programme, often under a constraint such as those posed by a given budget, or other considerations such as a given human-resource envelope. In a recent review of 23 papers based on 14 HIV optimisation models, the most common optimisation target, or optimand, was minimising HIV incidence, followed by maximising survival and utility (measured as disability-adjusted life-years (DALYs) averted or quality-adjusted life-years gained), minimising deaths or minimising costs (Avanceña et al., 2020).

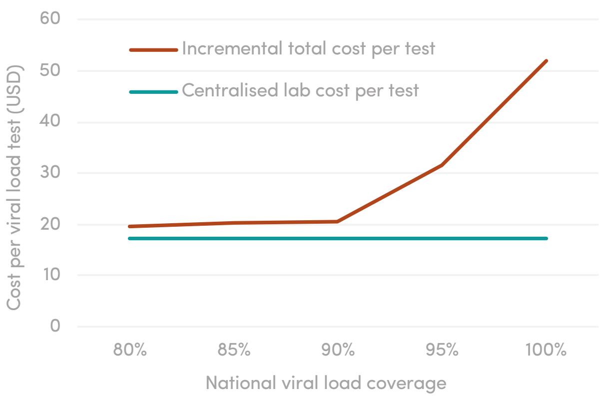
Other types of optimisation include spatial and temporal optimisation, as well as optimisation under explicit constraints such as health-systems capacity. Anderson et al. (2014) showed that targeting prevention interventions by population and locale was more impactful than a non-targeted approach under a limited budget in Kenya, with the locally focused approach reducing HIV infections by 33 percent per year compared with the blanket national-level approach. Kedziora et al. (2019) found that geographically optimising budget allocation within different regions in Ukraine would reduce DALYs lost to HIV by 26 percent, while optimisation across regions would reduce DALYs by 29 percent. Geospatial optimisation has its limits, however: Nichols et al. (2019) optimised the logistics of the sample-collection network for viral load tests in Zambia in order to increase test coverage from 80 to 100 percent of patient volumes (Figure 16.1) and found that, despite optimised sample-transport networks, cost per test would have to increase 2.6-fold due to the increased reliance on decentralised transport systems (Figure 16.2).

Figure 16.1: Maps of Zambian viral load sample transport networks at 80% (A) and 100% (B) of patient volumes



Source: Nichols et al., 2019. Reproduced with author's permission.

Figure 16.2: Cost per viral load test



Source: Nichols et al., 2019

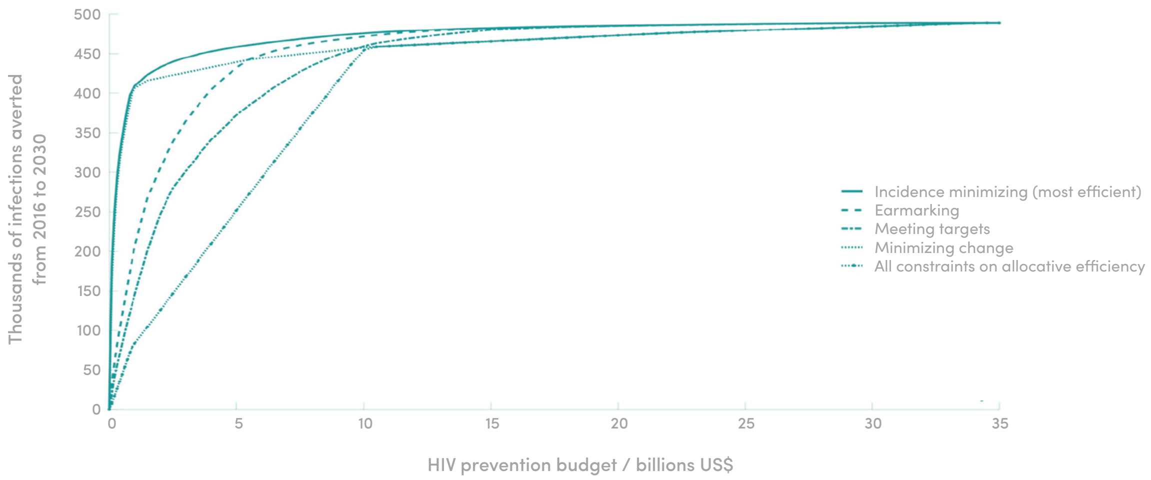
As an illustration of the impact of varying programme optimisation over time using the Optima HIV allocative optimisation model, Shattock et al. (2016) allowed annual HIV budgets for Zambia to vary over time within the same overall 10-year budget envelope and achieved a 7.6% decrease in projected new HIV infections compared with a constant baseline budget.

Combining different optimisation targets

There can be a tension between cost-effectiveness and other optimisation targets such as equitable access or ending AIDS as a public health threat.

Efficiency, allocative or otherwise, might not be the only, or not the primary, goal in allocating HIV budgets. However, tensions can arise if additional optimands are in juxtaposition to the main aim of allocative efficiency. For example, both internal and external healthcare targets can skew the HIV response away from its optimal impact, especially when resources for the implementation of these targets are constrained. This is especially severe when targets are set by international organisations that do not contribute major funding to their implementation. Stopard et al. (2019) optimised the allocation of HIV prevention budgets in Benin, Tanzania and South Africa considering a) the earmarking of funding to novel interventions, b) the attainment of the UNAIDS 90-90-90 targets and c) the stickiness of local planning processes. They found that all three reduced the impact of prevention programmes, though they only did so at budget levels lower than the current prevention budget (Figure 16.3). At higher budget levels, the impact of these technical inefficiencies became negligible.

Figure 16.3: Changes in HIV prevention efficiency under different scenarios and budget levels



Source: Stopard et al., 2019. Reproduced with author's permission.

Another tension arises when optimal allocation means the down-scaling of some interventions. Traditionally, HIV budgets in many countries have funded interventions such as the improvement of systems for patient tracking and documentation, supply chain management or pharmacovigilance (often summarised as programme enablers), which benefit health systems beyond HIV, or social enablers such as stigma reduction, community mobilisation and political commitment and advocacy work that prepare society's response to HIV more broadly. Where HIV funding has grown more than general development assistance, sometimes this has extended to programmes furthering other development aims (also called development synergies) such as social protection, education, legal reform, gender equality and the reduction of poverty and of violence by men against women (Schwartländer et al., 2011). These structural enablers and development synergies might need to continue to be funded in order to progress towards these other development aims, even though they might not contribute directly to the attainment of HIV endpoints or are less cost-effective in doing so than other HIV interventions.

Lastly, an additional optimand might be equitable coverage. Aiming for the highest coverage levels necessarily includes reaching underserved population groups, which in turn can increase both the average and marginal cost of an intervention so much so that it is no longer the most cost-effective. However, extending coverage to key populations at highest risk can also be cost-effective, such as in an analysis of different scenarios of expanding the HIV programme of

Côte d'Ivoire to allow different key population groups to reach the UNAIDS 90-90-90 targets by 2020 (and the 95-95-95 targets by 2030) alongside the general population. While the authors found that the maintenance of current coverage trends was almost three times more cost-effective than achieving the UNAIDS targets, they also found that a number of scenarios in which key population groups reached the UNAIDS targets first were more cost-effective than having all population groups reach these targets at the same time (Maheu-Giroux et al., 2019).

Additional aspects such as policymakers' obligations to reaching international goals or financial protection might play a role (Avanceña et al., 2020). Analytical methods such as multi-criteria decision analysis allow the consideration of these additional aspects, the elicitation of decision-makers' preferences between these additional criteria, and their weighting relative to each other, in order to produce rankings of interventions bearing all criteria out (Baltussen et al., 2016). Extended cost-effectiveness analysis additionally allows the joint consideration of the health and financial consequences, including financial risk protection and distributional benefits (Verguet et al., 2016). Despite these developments, there will remain a grey area of decision-making where alternative criteria are not known at the time of analysis, or cannot be quantified sufficiently.

Complexities in allocative efficiency modelling

Analysis that hopes to inform healthcare decision-making is complicated by (a) declining returns to investment in individual interventions, (b) increasing marginal costs at very high coverage levels, (c) interactions in effectiveness between different interventions and (d) lack of effectiveness data for some interventions.

Most countries' HIV programmes consist of a number of interdependent interventions. Once these have been implemented for a period of time and have reached high coverage levels, two interlinked analytical problems emerge. First, interaction effects increase as the number of interventions already implemented or considered for implementation increases. For example, scaling up any prevention intervention will likely reduce the need for treatment in later years, while scaling up treatment will reduce population HIV viral load and, by thus decreasing HIV transmission, will diminish the impact of prevention interventions (Chiu et al., 2017). Second, the relationship between the marginal cost of producing the next unit of output of an intervention and coverage is non-linear: average costs, in particular at the facility level, often decline initially with scale-up due to economies of scale but may increase at higher coverage levels because the remaining population groups tend to be harder to reach. Increasingly, models accommodate this "portfolio approach" to analysis, which makes it possible to include all interventions that are currently implemented from the same budget, or that are under discussion for funding, and analyse the cost-effectiveness of each while taking into account the impact of changing the coverage of one intervention on the need for all others.

The results of portfolio models can be presented in two ways. First, interventions can be ordered in a league table by their incremental cost-effectiveness over the same baseline. This allows policymakers to compare all interventions at a glance and to allocate budget to the most cost-effective interventions until the budget envelope is exhausted, i.e., until the total cost of the combined interventions is more than the available budget. While this allows an easy incorporation of notions of affordability into the analysis, the interdependence of interventions is still excluded. Alternatively, the incremental cost-effectiveness of each intervention can be analysed over a baseline including all more cost-effective interventions, incorporating diminishing returns. These are relevant in particular when considering HIV testing, whose yield of a newly diagnosed person living with HIV (PLHIV) must decrease over time as most PLHIV have already been diagnosed, and for prevention interventions whose returns decrease with decreasing incidence, at which stage more targeted approaches might be more useful. As Figure 16.4 shows, considering diminishing marginal returns greatly increases the incremental cost-effectiveness ratios (ICERs) but leads to more realistic estimates of each intervention's impact.

Figure 16.4: Comparison of conventional league table and optimisation routine in South African HIV investment case

Conventional league table		Optimisation routine		% change in ICER*
Rank	*ICER ($/LYS)*	*Rank*	*ICER ($/LYS)*	
Condom availability	Cost saving	Condom availability	Cost saving	
Male medical circumcision	Cost saving	Male medical circumcision	Cost saving	
SBCC 1 (HCT in adolescents, reduction in MSP)	46	ART (Eligibility at 500 CD4 cells/microl)	109	14%
ART (Eligibility at 500 CD4 cells/microl)	96	PMTCT	142	7%
PMTCT	132	Infant testing at 6 weeks	248	20%
Universal ART	186	Universal ART	249	34%
Infant testing at 6 weeks	208	SBCC 1 (HCT in adolescents, reduction in MSP)	749	1525%
HCT for sex workers	366	SBCC 2 (condoms)	1 200	112%
SBCC 2 (condoms)	566	General population HCT	1 236	-3%
SBCC 3 (condoms, HCT, MMC)	697	SBCC 3 (condoms, HCT, MMC)	1 816	161%
PrEP for sex workers	926	HCT for sex workers	2 643	621%
General population HCT	1 273	Infant testing at birth	2 937	118%
Infant testing at birth	1 349	PrEP for sex workers	9 947	974%
HCT for adolescents	1 772	HCT for adolescents	19 540	1003%
PrEP for young women	3 703	PrEP for young women Max	26 375	612%
Early infant male circumcision	8 712 984	Early infant male circumcision	89 642 731	929%

Source: Chiu et al., 2017. ICER: incremental cost-effectiveness ratio. $: US dollars. LYS: life-year saved. HCT: HIV counselling and testing. MSP: multiple sexual partners. PMTCT: prevention of mother-to-child transmission. ART: antiretroviral therapy. SBCC: social and behaviour change communication. MMC: medical male circumcision. PrEP: pre-exposure prophylaxis.

* Between methods

Lastly, some interventions, including a number of programme or other structural enablers, but also interventions implemented as packages, do not easily lend themselves to inclusion in optimisation models, as their effectiveness has often not been measured at all, or has not been measured against HIV-relevant endpoints such as incidence, reduction of AIDS-related deaths, or uptake of other interventions with known effectiveness. If there are reasons to believe that structural enablers provide a role in increasing uptake or demand for other interventions, it is often still possible to include their cost alone, based on programme data regarding the types and quantities of resources needed for their implementation; but the lack of effectiveness data prevents an assessment of their impact on a country's HIV epidemic in a transmission model and, hence, ultimately their cost-effectiveness.

Trade-offs over time

Trade-offs also apply over time. Increasing intervention coverage today improves health outcomes but also affects spending needs in the future.

As with many other diseases, decisions regarding the funding of HIV programmes must take into account the timing of expenditure. In HIV, it is additionally important to consider the timing of the impacts of interventions which often come to fruition quite a bit later than the upfront investment – in particular in prevention interventions. Interventions such as ART have early and sustained impacts on survival and transmission but create a longer-term financial commitment that has to be factored into the initial decision-making process. This is particularly important when motivating funding towards longer-term aims such as "ending AIDS" as a public health threat. UNAIDS' analyses as part of the "Fast-track" programme launched in 2014 projected that front-loading the investment would save US$ 24 billion of costs for HIV treatment annually by 2030 – while also saving 21 million lives and preventing 28 million HIV infections (UNAIDS, 2014).

Modelled analyses additionally aim to choose a time horizon that incorporates the bulk of both costs and outcomes of the intervention under analysis, including the impact on onward transmission and secondary cases averted, and future births and deaths. Throughout the analyses mentioned above, the costs of countries' HIV programmes are projected to increase over the first years, before they fall again – at least for interventions whose coverage is sub-optimal at baseline. Care needs to be taken, then, in choosing the correct time horizon. The time-varying optimisation by Shattock et al. (2106) mentioned above found that recommendations regarding which interventions to prioritise changed greatly between analyses with a 5-, 10- and 20-year programme horizon, depending on the timing of costs and benefits and the discount rates applied; a finding supported by Haacker et al. (2020).

Another important aspect in HIV decision-making is framing the decision problem so that the cost of inaction can be incorporated. In many analyses this is done in the shape of a baseline or counterfactual representing the current HIV programme (by keeping the choice of interventions and their coverage constant), or, in early analyses, the counterfactual of "no HIV care". While these scenarios are often difficult to estimate, one analysis of the HIV programme in Zimbabwe using the Optima model found that a scenario of "no funding" would increase HIV infections by between 80,000 and 120,000 annually over 15 years and lead to between 30,000 and 100,000 HIV-related deaths, compared to below 20,000 infections and deaths per year expected under a baseline of keeping the current programme constant (World Bank, 2019). In situations of decreasing funding or traditional funders disengaging altogether, more pessimistic counterfactuals might be warranted.

Practical and political limitations of allocative optimisation models

While adding these optimisation targets into analyses often produces impressive results vis-à-vis the status quo, a number of ethical and political aspects apply that might reduce their feasibility and their ability to be implemented. Among these are the political and practical issues associated with focusing resources on certain areas of a country (Anderson et al., 2014; Meyer-Rath et al., 2018) or ignoring district boundaries in planning and budgeting (Nichols et al., 2019), the non-fungibility of resources which does not allow for quick "switching" on and off of interventions over time (Shattock et al., 2016), the influence of the targets of donors and international organisations (Avanceña et al., 2020), and the presence of constraints on the supply and demand side that are unknown or hard to quantify at the time of analysis (Vassall et al., 2016).

However, adding these additional aspects into an optimisation model holds two risks. First, the requirement for additional data representing each of these aspects, across all geographic and population levels at which programmes could be optimised, limits the number of countries that can plausibly hope to gain useful insights from these models. Second, over-interpretation of the models' results, given the level of uncertainty in their inputs, might in turn lead to net-injurious programming, for example if the wrong sub-populations or geographical areas are targeted for prevention interventions, or targeting of incidence "hotspots" jeopardises general population coverage. These aspects might reduce the role of optimisation models in planning HIV programmes.

References

- Anderson SJ, Cherutich P, Kilonzo N, Cremin I, Fecht D, Kimanga D, Harper M, Masha RL, Ngongo PB, Maina W, Dybul M, Hallett TB. 2014. "Maximising the Effect of Combination HIV Prevention through Prioritisation of the People and Places in Greatest Need: A Modelling Study". *Lancet*, vol. 384, no. 9939, pp. 249-56.
- Avanceña ALV, Hutton DW. 2020. "Optimization Models for HIV/AIDS Resource Allocation: A Systematic Review". *Value Health*, vol. 23, no. 11, pp. 1509-21.
- Baltussen R, Niessen L. 2006. "Priority Setting of Health Interventions: The Need for Multi-criteria Decision Analysis". *Cost Effectiveness Resource Allocation,* vol. 4:14.
- Chiu C, Johnson LF, Jamieson L, Larson BA, Meyer-Rath G. 2017. "Designing an Optimal HIV Programme for South Africa: Does the Optimal Package Change When Diminishing Returns are Considered?" *BMC Public Health*, vol. 17:143.
- Haacker M, Hallett TB, Atun R. 2020. "On Time Horizons in Health Economic Evaluations". *Health Policy and Planning*, vol. 35, no. 9, pp. 1237–1243.
- Kedziora DJ, Stuart RM, Pearson J, Latypov A, Dierst-Davies R, Duda M, Avaliani N, Wilson DP, Kerr CC. 2019. "Optimal Allocation of HIV Resources among Geographical Regions". *BMC Public Health,* vol. 19:1.
- Maheu-Giroux M, Diabaté S, Boily MC, Jean-Paul N, Vesga JF, Baral S, Abo K, Wognin V, Diouf D, Alary M. 2019. "Cost-Effectiveness of Accelerated HIV Response Scenarios in Côte d'Ivoire". *Journal of Acquired Immune Deficiency Syndromes*, vol. 80, no. 5, pp. 503-12.
- Meyer-Rath G, McGillen J, Cuadros D, Hallett T, Bhatt S, Wabiri N, Tanser F, Rehle T. 2018. "Targeting the Right Interventions to the Right People and Places: The role of Geospatial Analysis in HIV Programme Planning". *AIDS*, vol. 32, no. 8, pp. 957-63.
- Nichols BE, Girdwood SJ, Crompton T, Stewart-Isherwood L, Berrie L, Chimhamhiwa D, Moyo C, Kuehnle J, Stevens W, Rosen S for EQUIP Health. 2019. "Monitoring Viral Load for the Last Mile: What Will It Cost?" *Journal of the International AIDS Society,* vol. 22:e25337.
- Schwartländer B, Stover J, Hallett T, Atun R, Avila C, Gouws E, Bartos M, Ghys PD, Opuni M, Barr D, Alsallaq R, Bollinger L, de Freitas M, Garnett G, Holmes C, Legins K, Pillay Y, Stanciole AE, McClure C, Hirnschall G, Laga M, Padian N. 2011. "Towards an Improved Investment Approach for an Effective Response to HIV/AIDS". *The Lancet,* vol. 377, no. 9782, pp. 2031-41.
- Shattock AJ, Kerr CC, Stuart RM, Masaki E, Fraser N, Benedikt C, Gorgens M, Wilson DP, Gray RT. 2016. In the Interests of Time: Improving HIV Allocative Efficiency Modelling via Optimal Time-varying Allocations". *Journal of the International AIDS Society,* vol.19:20627.
- Stopard IJ, McGillen JB, Hauck K, Hallett TB. 2019. "The Influence of Constraints on the Efficient Allocation of Resources for HIV Prevention". *AIDS*, vol. 33, no.7, pp. 1241-46.
- UNAIDS (Joint United Nations Programme on HIV/AIDS). 2014. *Fast-Track: Ending the AIDS Epidemic by 2030*. Geneva: UNAIDS.
- Vassall A, Mangham-Jefferies L, Gomez GB, Pitt C, Foster N. 2016. "Incorporating Demand and Supply Constraints into Economic Evaluations in Low-income and Middle-income Countries". *Health Economics,* vol. 25, supplement 1, pp. 95-115.
- Verguet S, Kim JJ, Jamison DT. 2016. "Extended Cost-Effectiveness Analysis for Health Policy Assessment: A Tutorial". *Pharmacoeconomics*, vol. 34, no. 9, pp. 913-23.
- World Bank. 2019. *Phase 2 Report, Improving the Allocative Efficiency of the HIV Response Across the Care Cascade In Zimbabwe: Findings from a Modelling Analysis*. Washington, DC: World Bank.