



PEPFAR

U.S. President's Emergency Plan for AIDS Relief



USAID

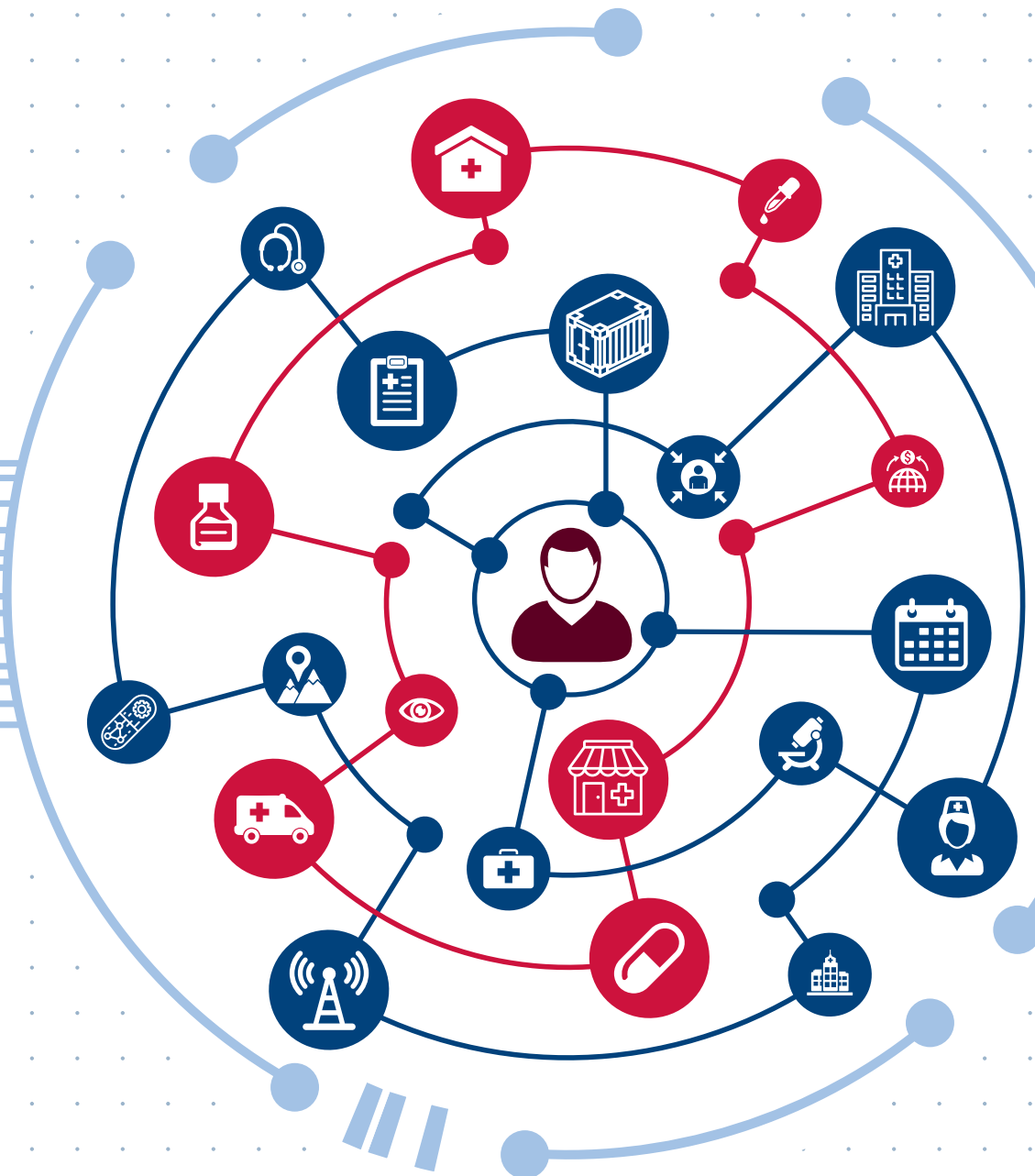
FROM THE AMERICAN PEOPLE

**CENTER FOR
INNOVATION
AND IMPACT**

USAID | Global Health

DATA AND ADVANCED ANALYTICS IN HIV SERVICE DELIVERY

Use Cases to Help Reach 95-95-95



The lead author of this report was Boston Consulting Group (BCG), in collaboration with the United States Agency for International Development (USAID) South Africa with additional funding and support from USAID's Center for Accelerating Innovation and Impact (CII). This report could not have been written without the support and invaluable contributions of the USAID team, programme partners, industry experts, workshop participants, and all who contributed their time, insights and experience. The authors particularly wish to thank the participants of this review from Project Last Mile and Panagora Group.



Launched in 2003, the U.S. President's Emergency Plan for AIDS Relief (PEPFAR) is the largest commitment by any nation to address a single disease in history, enabled by strong bipartisan support across nine U.S. congresses and three presidential administrations, and through the American people's generosity. The USAID Office of HIV/AIDS drives the Agency's efforts under PEPFAR in controlling the HIV/AIDS epidemic by providing global expertise to strengthen the impact of USAID's overall response.



USAID's Center for Innovation and Impact (CII) takes a business-minded approach to fast-tracking the development, introduction, and scale-up of health interventions that address the world's most important health challenges. CII invests seed capital in the most promising ideas and novel approaches, using forward-looking business practices to cut the time it takes to transform discoveries in the lab to impact on the ground.



Boston Consulting Group (BCG) is a global management consulting firm and a leading advisor on business strategy. BCG works with organizations in the social, public, and private sectors to deliver solutions to the most complex problems in contemporary society.



BCG GAMMA is BCG's global team dedicated to applying artificial intelligence and advanced analytics to business at leading companies and organizations. The team includes 800-plus data scientists and engineers who apply AI and advanced analytics expertise (e.g., machine learning, deep learning, optimization, simulation, text and image analytics) to build solutions that transform business performance. BCG GAMMA's approach builds value and competitive advantage at the intersection of data science, technology, people, business expertise, processes and ways of working.

For contact information, and to download the latest version of *Data and Advanced Analytics in HIV Service Delivery: Use Cases to Help Reach 95-95-95*, please visit www.usaid.gov/cii.

Table of Contents

1. Foreword.....	2
2. Introduction.....	3
3. Executive Summary.....	5
4. Understanding pain points.....	7
4.1 Key themes	7
4.1.1 Focusing.....	7
4.1.2 Tailoring.....	8
4.1.3 Touchpoints.....	8
4.1.4 Tracking.....	9
4.1.5 Responsiveness.....	9
5. Addressing pain points through use cases.....	11
5.1 Introduction to advanced analytics use cases	11
5.2 Prioritization criteria	11
5.3 Ten impact-generating use cases	12
5.4 Key learnings.....	21
5.4.1 Factors affecting impact.....	21
5.4.2 Feasibility.....	23
6. Data sources, challenges, and opportunities.....	25
6.1 Patient-level data in LMIC countries	25
6.2 External data considerations and sources	27
7. Next steps: Key considerations for developing a use case	30
8. Conclusion	33
9. Appendix	34
A. Glossary.....	34
B. Long list of pain points	35
C. Long list of use cases	40
D. Detailed descriptions of impact and feasibility sub-criteria	42
E. Categories and details of external data sources	43

I. Foreword

I write this from my post as a member of the United States Agency for International Development (USAID) team in South Africa and a privileged contributor to this compendium of good ideas. The United States government has had the privilege of supporting the South African government to help maintain almost 4 million people living with HIV on treatment, with the aim of supporting all 7.5 million people living with HIV before the end of 2020. With an ever-changing and complex world, I am grateful to the team at Boston Consulting Group's (BCG) Scale for Impact project for pulling together this useful tool when the world needs it most.

South Africa made strides in controlling the HIV epidemic by developing some of the most innovative approaches to differentiating models of treatment delivery. While the HIV program is no stranger to combatting fear, stigma, limited resources, and staff shortages, the health system will be tested more than ever in light of COVID-19. The disease entered South Africa's borders as the country was making tremendous strides toward reaching its global 90-90-90 HIV epidemic control targets. The introduction of COVID-19 presents new challenges for the provision of care and calls for innovative ways to address the barriers that clients face when obtaining care, so that we don't lose the gains of recent years.

Despite its focus on data and advanced analytics, this report is above all else about the client and the health worker. The authors mapped the common journeys of clients (typified by the character named "Gift") and health workers (typified by the character "Ruth"). In a report about data, it's crucial to humanize these groups and provide narratives of their experiences to illustrate pain points faced by the clients and staff. Pain points are the millions of tiny inconveniences and instances of confusion, duplication, and worry that prevent large-scale programs from sustainably succeeding. This report displays a comprehensive tabulation of what a common HIV program faces in providing a framework for readers to prioritize the issues in their own contexts. Print out the accompanying graphic showing the synthesized list of pain points, hang it over your computer, look at it every day—never forget the human factors that drive epidemic control and an efficiently operating program.

The authors of this report also gifted us all with a modern, shiny toolbox: a range of approaches related to predictive modeling, machine learning, artificial intelligence, sentiment analysis, and geospatial analytics, tied under the common theme of advanced analytics. For each problem, the authors tapped into one of the largest teams of data scientists in the world, the BCG GAMMA team, to identify problem-solving use cases that could be applied using the data commonly available in a national HIV program and modern data sources.

To the reader, may the ideas in this report inspire you to ask, "How might we...?" How might we use data from bus routes and satellite maps to deliver services to clients who move for seasonal work? How might we promote client adherence by optimizing how we implement appointment reservations at facilities? How might we use client attributes to learn who may be at higher risk for getting lost in the health system? How might we achieve a balance of preserving ethical protections for individual data while leveraging the possibilities of modern technology? These questions and more are reviewed in this report with recommendations for resourcing and implementation.

Many of us working in HIV programs globally are often enmeshed in our own program data. This report reminds us that there are diverse sources of information relevant to our efforts beyond the conventional. After all, the clients we serve often use social media, take public transportation, seek private health care, and buy goods from local pharmacies. The application of diverse data sets emphasizes the importance of clients' lives outside of health facilities. When we design programs that synthesize the complex lives of our clients and health workers, we affirm their personhood beyond merely seeing them as patients and clinicians. This is a necessary step toward controlling the HIV epidemic and ensuring the global effort to control COVID-19 better prepares the global public health community to protect the well-being of future generations.

Josef Tayag
Health Systems Strengthening Team Lead – Acting
USAID South Africa/Health Office

2. Introduction

Global health efforts have made significant progress against the HIV/AIDS pandemic.¹ At the turn of the millennium, less than seven hundred thousand of the almost thirty four million people living with HIV (PLHIV) had access to antiretroviral therapy (ART). In 2019, that number stood at around twenty-five million of the thirty eight million PLHIV.

As the gap to reaching targets of controlling and eliminating this pandemic closes, the difficulty to achieve them increases. The year 2020 marks a transition: the Joint United Nations Programme on HIV/AIDS's (UNAIDS) target of 90-90-90² by the end of this year becomes 95-95-95³ by 2030. However, many countries will struggle to achieve 95-95-95 in the next ten years.

Global data suggests that in 2018, 79% of PLHIV were aware of their status, and 78% of PLHIV aware of their status were on stable treatment, equivalent to 62% of all PLHIV. Of the PLHIV on stable treatment, 86% were virally suppressed, which is a laudable achievement, but only equivalent to 53% of all PLHIV.^{4,5} Achieving 95% treatment rates will require continued innovation of personalized and differentiated care models. General, widespread interventions—while important to maintain a critical mass—will not be enough for the remaining individuals who are hard to find, who have difficulty being linked to treatment, or who have dropped out of treatment.

We think that data could make the difference. Vast volumes of structured and unstructured data are collected on the millions of PLHIV globally, but sufficient insight isn't being gained from it. This could be improved through more widespread use of advanced analytics techniques such as clustering to better understand at-risk populations, predictive modeling to understand likelihood of risk for an individual, and optimization of ART inventory. The insight garnered using these techniques can be used to predict patterns of disease, optimize treatment decisions, and improve patient experience, to name a few examples.

Advanced analytics has an important role to play by addressing crucial pain points that contribute to the gap between where we are today and 95-95-95. In the private sector, advanced analytics has been leveraged for years with success ranging from optimizing manufacturing processes to understanding customers better to enable targeted product offerings. Identifying high-impact, feasible use cases for advanced analytics and addressing the data access and quality requirements to support them are interlinked and must be tackled together. Better data quality and integration in low- and middle-income countries (LMIC) could ensure the usefulness of this data. Without this improvement, it will be difficult to close the gap toward 95-95-95 in the years to come. This is not just an opportunity, but an imperative: advanced analytics is the lever enabling us to meet increasingly ambitious goals in a resource- constrained world. In countries where the HIV/AIDS burden is high, budgets are being stretched, both for global donors and local governments, even as targets to control HIV increase.

This report reflects the output of an analysis conducted by Boston Consulting Group (BCG) on behalf of the United States Agency for International Development (USAID). The objective of the analysis was to identify high-priority advanced analytics use cases that could improve HIV service delivery globally. The identification of use cases is necessarily informed by a review of the most pressing challenges in HIV service delivery, with a focus on improved data quality and utilization. Based on this review, our ideation process led to a long list of use cases that we prioritized and refined. We also considered enablers for these use cases: how the quality of patient-level treatment data can be improved and how external data sets can be better leveraged to enrich insight.⁶ Lastly, for those looking to apply analytics use cases in their own work, we speak to key considerations for implementing a portfolio of use cases over time.

Our objective is to help drive global thinking on how to leverage advanced analytics to improve HIV service

delivery, and more specifically, to support program managers and implementers to include use cases in their strategy to reach 95-95-95. Key learnings gained from

our analysis can provide a useful launch pad for others seeking to do similar work in different contexts.

Footnotes

1. Malaria No More and World Health Organization statistics.
2. This means 90% of PLHIV are aware of their status, of whom 90% are on HIV treatment, of whom 90% are virally suppressed.
3. This means 95% of PLHIV are aware of their status, of whom 95% are on HIV treatment, of whom 95% are virally suppressed.
4. Data from the UNAIDS treatment cascade, published in 2019.
5. Viral suppression refers to an HIV-positive person having a very low level of HIV in his or her body: less than 1,000 copies of HIV per milliliter of blood, according to UNAIDS.
6. For the purposes of this work, external data sets are defined as non-health data sets, and health data sets curated or owned within the private sector; that can provide insight into population behavior and area trends.

3. Executive Summary

Data and advanced analytics offer an opportunity to improve service delivery for the millions of PLHIV across the globe, while using limited resources more effectively. The first step toward harnessing analytics toward the ultimate goal of reaching an AIDS-free generation is understanding where issues arise in HIV service delivery along the patient journey, across diagnosis, treatment, and adherence. In this analysis, over two hundred pain points were gathered from stakeholder discussions, covering each step in the patient journey as well as three cross-cutting categories: data management, operations management, and staff. From our discussions, we saw a concentration of pain points in early treatment, which is understandable considering the challenges faced in keeping PLHIV on treatment in the first few months, when the majority of patient loss to follow-up (LTFU) usually occurs. Across the patient journey, pain points can be categorized into five key themes: Focusing, Tailoring, Touchpoints, Tracking, and Responsiveness, which must be addressed to achieve 95-95-95 in countries with a high HIV burden and are described in Section 4.

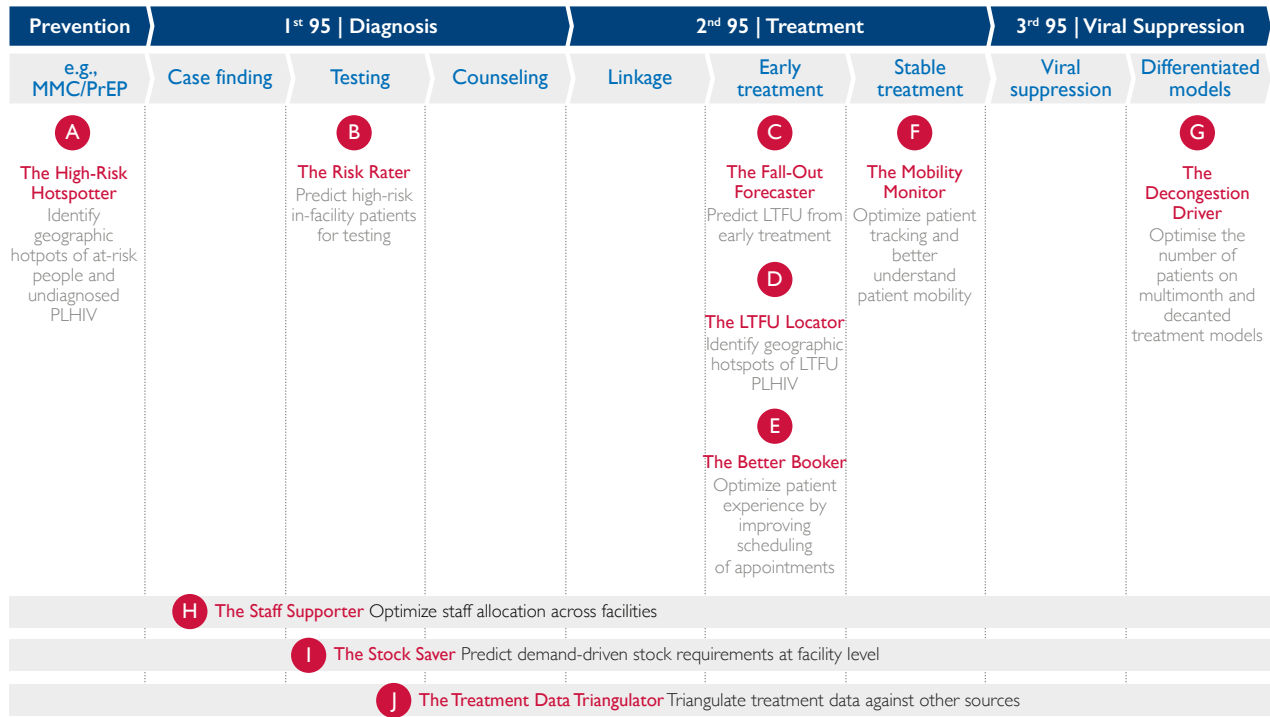
Advanced analytics use cases are proposed as a solution to address many of these patient journey pain points. Around forty use cases were conceptualized in a long list and prioritized by impact and feasibility. The resulting shortlist of ten priority use cases is shown in Figure 1.

Detailing the features and end-to-end processes of these ten use cases helped us to do a more in-depth impact and feasibility assessment of them. Although these use cases have global relevance for the HIV landscape, specific priorities could differ based on a country's context, constraints, and 95-95-95 objectives. A key example of this is whether a country's focus is on initiation or retention. Another growing focus area is retaining patients in less resource-intensive treatment models, such as decentralized medicine pick-ups.¹ The concentration of use cases around the early and stable treatment steps of the patient journey reflects this. Please see the long list of use cases in the appendix to this report for additional suggestions of relevant use cases for other treatment steps.

As the overviews of the use cases provided in Section 5.3 show, benefits to undertaking a use case vary but are projected to be significant across multiple dimensions of clinical outcomes improvement as well as cost reduction. For example, patient LTFU can be reduced by over 10% in some cases, and cost reductions of up to 15% in the relevant spend category can be achieved. Common feasibility challenges arose across the ten priority use cases relating mainly to change management and data requirements. Mitigating actions to these challenges are particularly important, such as reducing the number of end users for the use case, centralizing the use case and avoiding real-time processing, determining data requirements early on, partnering with the owners of patient-level data sets, and avoiding input data from on-the-ground collection wherever possible. Overall, more use cases in the top ten qualify as “big bets”—high impact but relatively lower feasibility—rather than “quick wins”—relatively higher feasibility but lower impact—which shows the exciting potential of advanced analytics use cases but the significant effort required to realize that potential.

All use cases require quality data, particularly at the patient level, and ensuring quality means ensuring timely and accurate capturing and validation of patient-level treatment data, establishing clearer ownership and role delineation, and creating streamlined programmatic tools to support these improvements, among other best practices described in Section 6.1. Data outside the public health ecosystem can also play an important role in providing insights into population behavior and trends, as well as area-level events that could affect the population and facility operations in an area. A wealth of population and area data is available, as described in Section 6.2 and the appendix, with data vendors offering purchase or subscription packages and research institutions providing open-access resources. In many instances, however, data partnerships with other organizations that either own or aggregate private-sector data are necessary to unlock the full potential of use cases. These partnerships require rigorous adherence to data privacy regulations, strict data governance processes, and a carefully considered approach toward establishing data-sharing agreements.

Figure 1: Ten impact-generating use cases along the patient journey



Where to from here? Although having a “portfolio” of priority use cases to choose from is helpful, most organizations will be able to roll out only one to two use cases successfully within a year. This is due to the intensity of the use case delivery model: the process of taking a use case from an idea to a fully deployed and scaled-up product in use by everyday operations. Each phase of the delivery model brings different considerations, but an overarching imperative is to ensure that the delivery squad, the team responsible for the use case development, is fully dedicated to make it a success. This requires strong governance, as well as dedicated capacity from the product owner, end-user representatives, and subject-matter experts. Using an agile, sprint-based approach can allow impact to be seen on the ground in an earlier stage in the process rather than after full rollout.

Throughout this report, you will be accompanied by two illustrative personas, Gift and Ruth. Gift is a young man who was recently diagnosed with HIV, and Ruth is a case manager working in a large public hospital.² Their pain points are highlighted as examples of first-hand

experiences with challenges in the HIV landscape. We also use instances in their daily lives to show what HIV service delivery could look like with a few of our highest-impact use cases implemented. Although we provide detail to illustrate our analysis, please see the appendix for the long lists of pain points and use cases referred to in the report.

A user-centric, solution-driven approach has been adopted to build use cases bottom-up that address the specific issues faced by stakeholders in the HIV landscape. This approach also ensures that, once implemented, the solutions have strong user pull. We chose not to focus on creation of digital tools or technologies, such as apps, websites, block chain, and the Internet of Things, but rather we focus the use case design on describing what is the actual data insight or usage that creates value for the system. Therefore, in some cases, an app or other platform may serve as the interface for how a user interacts with the use case, but it is not the focus of the use case. This definition was relevant for our work, but may be too restrictive for different contexts, constraints, and objectives.

Footnotes

1. Decentralized models provide patients with a channel to receive ART outside a health care facility. Please see 4.1.3, “Touchpoints,” for more explanation.
2. Please note that these personas are entirely fictional, and any resemblance to actual persons is coincidental.

4. Understanding pain points

To understand the potential impact of advanced analytics use cases in HIV service delivery, it is important to determine where data-driven interventions are the most necessary. The patient journey, illustrated in Figure 2, describes the different stages of HIV care and treatment for patients who remain on ART until they are virally suppressed and transferred to differentiated models.¹ Although the illustration is too linear to be an accurate representation of the varied and often interrupted journeys that PLHIV take, it helps to differentiate events and stages at which use cases can be targeted to address inefficiency. In addition, the patient journey is a globally relevant representation of steps toward 95-95-95.

To identify pain points, stakeholders in the HIV landscape were asked about the biggest challenges in their day-to-day work—the unanswered questions or missing information that, if addressed, could have the largest impact on achieving 95-95-95.^{2,3,4} The pain points heard were organized by the patient journey steps to understand where and how the pain points affect HIV service delivery. To properly capture cross-cutting pain points, three overarching categories were included: data management, operations management, and staff.

Pain points heard were most concentrated in early treatment. This is to be expected given the challenges PLHIV face when starting an intensive treatment regimen. However, pain points will differ based on where a country experiences the greatest gap to achieving 95-95-95. For example, in some sub-Saharan African countries, the greatest gap is in the first 95: diagnosis. In 2018, the share of PLHIV who knew their status was 72%

in Mozambique, 42% in Angola, and only 24% in South Sudan.⁵ In some contexts, up to 80% of treatment LTFU is estimated to occur in the first six months of treatment due to lack of overall case management, side effects of ART, poor initial health, inconvenient treatment options/hours, and poor professional service or judgment from health care workers (HCWs).⁶

4.1 Key themes

Five key themes emerged from our pain point discussions: Focusing, Tailoring, Touchpoints, Tracking, and Responsiveness (see Figure 3).

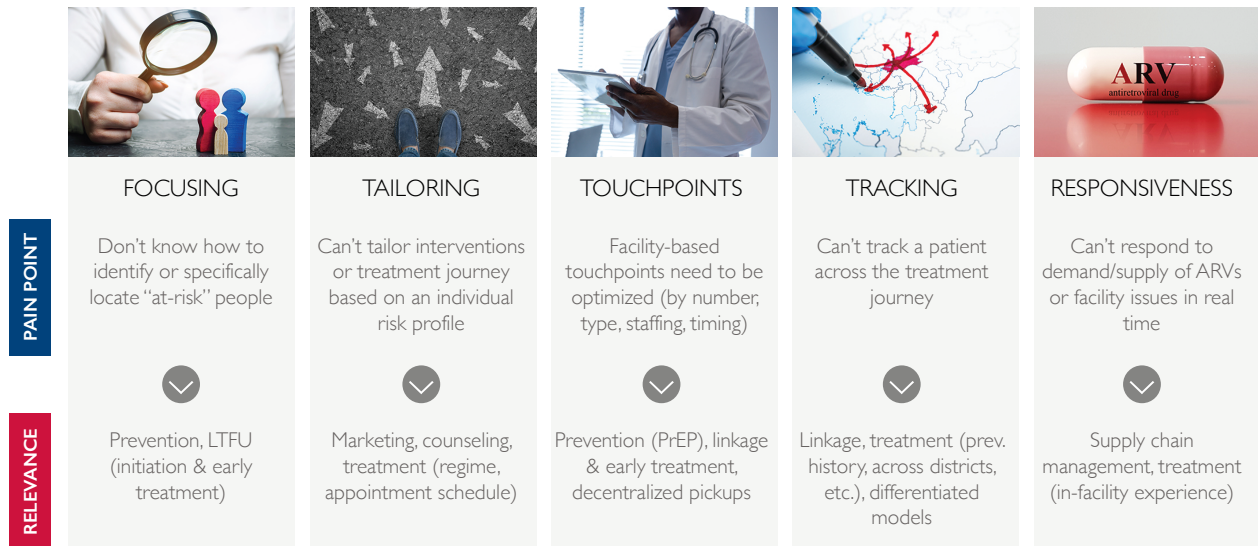
4.1.1 Focusing

Focusing refers to identifying and locating those at high risk of being HIV positive. In sub-Saharan African countries with high incidence rates outside of key populations,⁷ case finding has become increasingly complex. Program managers face difficulty in segmenting vulnerable populations,⁸ which are comparatively large, and whose high-risk individuals are often not easily identifiable. Information that could be used to identify high-risk individuals is often sensitive and hard to come by, such as their sexual history and incidence of sexually transmitted infection. Similar concerns apply to undiagnosed and LTFU PLHIV. Although we have a general understanding of risk factors, we struggle to identify risk profiles based on individuals' unique combination of these factors. In addition, it is difficult to locate high-risk individuals and communities. If we could better understand risk and identify risk pockets, we could better focus resources where they are most needed.

Figure 2: Linear illustration of a patient’s journey through HIV care and treatment



Figure 3: Key themes of pain points heard



4.1.2 Tailoring


Tailoring refers to personalizing interventions to reach those at risk, or PLHIV, in the best and most effective way possible. A recurring theme from our discussions was the need for tailored interventions that consider risk factors, past behavior, and future preferences, as the current "one-size-fits-all" approach can contribute to treatment dropout for some patients. Personalized interventions could be as simple as matching male counselors to male patients, or ensuring that experienced and high-performing case managers are allocated the patients at highest risk of LTFU. Tailoring could also be done at a program level, for campaign planning and marketing, to ensure that key messages are tailored to the correct communities and the most pressing issues. This would go a long way to ensuring that the needs of the patient, as the ultimate client of the health care system, are met.

4.1.3 Touchpoints

Touchpoints refers to the number of in-facility appointments required for treatment. Touchpoints' frequency was raised as a contributing factor to treatment dropout across the patient journey. A person on pre-exposure prophylaxis (PrEP) must regularly visit a health care facility for a prescription and liver function tests. A patient on ART must visit the facility several times in the first few months of treatment for various appointments. Even once in differentiated treatment models, such as multimonth dispensing, adherence clubs, or ART pick-ups outside facilities, patients still need to return to a health care facility multiple times per year for clinical checkups. This frequency can be a significant driver of dropout.

Repeated visits across the patient journey can be inconvenient and difficult to navigate, particularly

e.g., MMC/PrEP
Case finding
Testing
Counseling
Linkage
Early treatment
Stable treatment
Viral suppression
Differentiated models



A DAY IN THE LIFE: GIFT

Gift has just turned twenty and is struggling to come to terms with his diagnosis. He visits the clinic between his home and place of work early on a Tuesday morning, hoping not to run into any of his mother's friends, who often bring their children to these types of clinics. When Gift eventually enters the facility, he remembers being told about dedicated areas for men to receive treatment and counseling, but he can't see any sign and doesn't want to ask the nurse, who is already looking at him strangely.



A DAY IN THE LIFE: GIFT

Gift works in a call center. It's a new, exciting job that could provide him with further opportunities. He has not told his manager about his recent diagnosis and doesn't plan to. However, Gift's late arrival to work a few times in the last few months has prompted a negative review from his manager, who isn't buying the excuse of transport issues.

for those employed, and can also lead to stigma. Compounding the number of touchpoints is their duration: waiting times in clinics in LMICs can be several hours. Reducing extensive in-facility waiting time is important to improve patient satisfaction and retention. Although strong arguments exist for regular patient checkups to confirm adherence and manage side effects, checkups could be planned more effectively for individual patients, or better support could be given to patients over facility-intensive periods. From a patient perspective, pain points in treatment touchpoints can cascade to affect many other aspects of a patient's personhood.

4.1.4 Tracking

Tracking refers to being able to follow a patient's data record across the patient journey, to ensure the patient stays adherent and does not become LTFU. Without intensive and cooperative case management, transitioning patients often "fall out" of databases and get marked as LTFU. A similar issue arises when PLHIV move geographically across facilities, districts, or provinces. A patient moving, even temporarily, can become lost in the system. Failure to account for this

movement, referred to as "silent transfer," could make up as much as one-third of total recorded LTFU. A more accurate understanding of retention rates could help identify where the real issues are and reduce HCWs' wasted effort to track perceived LTFU patients who are in fact still on treatment. Pain points were also raised around the lack of understanding of how population movements, driven by seasonal factors and migration from other countries, affect program planning.

4.1.5 Responsiveness

Responsiveness refers to being able to respond to issues in HIV service delivery in a quick and effective data-driven way. Rich data on issues facing communities, facilities, or larger areas is often not captured and input effectively into decision-making to improve HIV service delivery. This is particularly relevant to supply chain management. An oft-cited pain point by those working with antiretroviral (ARV) supplies is the lack of demand information to help predict stock requirements in advance. Multiple factors can significantly change demand for ARVs, including cross-border stockpiling of drugs by migrants from poorer neighboring countries, seasonal patient movements,



A DAY IN THE LIFE: RUTH

Something is bothering Ruth. A patient of hers, Ahmad, visited a different area of the country a month ago to care for his sick mother and has not returned. Ahmad assured Ruth that he would be back in time for his appointment, but days have passed and there is no sign of him. Ruth has left messages on Ahmad's phone but has had no response. Ruth doesn't know where to start looking. She's worried that Ahmad has dropped off treatment. Ahmad has hypertension, and it's important that he stay adherent to his medication.

private-to-public sector crossover, and changing prescriptions. Having a holistic, up-to-date view of these factors—and how they might change looking forward—

could help limit the significant patient satisfaction and financial cost of under- and overstocking ARVs.

Footnotes

1. Decentralized models provide patients with a channel to receive ART outside a health care facility. Please see 4.1.3, "Touchpoints," for more details.
2. We did not address specific problems (e.g., treatment dropout) exhaustively. Rather, we focused on the operational issues that create inefficiency across HIV service delivery.
3. Stakeholders included funders, program managers, supply chain managers, care provider representatives, and civil society representatives.
4. Over two hundred pain points were garnered from these conversations. A list of these can be found in the appendix to this report.
5. UNAIDS data, 2019.
6. Johnson LF, et al., "Estimating the impact of antiretroviral treatment on adult mortality trends in South Africa: A mathematical modelling study," Progress Report: Operation Phuthuma, April 24 2019, using proportion of LTFU across time periods for the Jul–Sep 17 cohort.
7. Key populations are defined as men who have sex with men, transgender people, sex workers, and people who inject drugs.
8. Vulnerable populations are defined as adolescent girls and young women.

5. Addressing pain points through use cases

5.1 Introduction to advanced analytics use cases

Based on a clear view of the pain points in HIV service delivery, we ideated a long list of advanced analytics use cases. To ensure ideas are workable and creating value, we defined a use case according to the following criteria:

Being data-driven: Data must be a key input to the use case, which should either use historic data to generate insights, or optimize data collection or ingestion to directly enable the quicker and easier derivation of insight. Information technology (IT) infrastructure or a data tool that would not improve insight derivation was excluded.

Using advanced analytics: The use case should leverage one or more advanced analytics techniques, including machine learning, predictive modeling, optimization, and geo-analytics (see Figure 4). We did not limit our definition to artificial intelligence (AI).

Having measurable impact: The impact of a use case should be clearly measurable as it relates to the achievement of 95-95-95, the optimization of spending or improved efficiency of a process, and/or the direct improvement of satisfaction levels for players in the value chain.

Having specified users: The primary stakeholder(s) that will use and benefit from the use case must be clearly defined.

Being used continuously: The use case must be able to adapt to new input data and generate insights or make predictions. Once-off analysis for research purposes or static insights, while valuable to programs, was excluded to ensure the use cases had ongoing benefit through improvement to daily ways of working.

Around forty use cases were developed to address the pain points discussed in the previous section. Please see the appendix to this report for the long list of use cases and more information on the methodology used.

5.2 Prioritisation criteria

Taking even one use case from an idea to full deployment is a complex process requiring significant time, effort, and resources. Prioritizing our long list of use cases was essential to determine which ones were the most worthwhile to discuss further.

Prioritization was based on an assessment of impact, defined as the improvement to HIV service delivery, and feasibility, defined as the ease of deploying a use case at scale. To ensure that the most important aspects were taken into consideration, we defined sub-criteria for both impact and feasibility.

Impact criteria included the following:

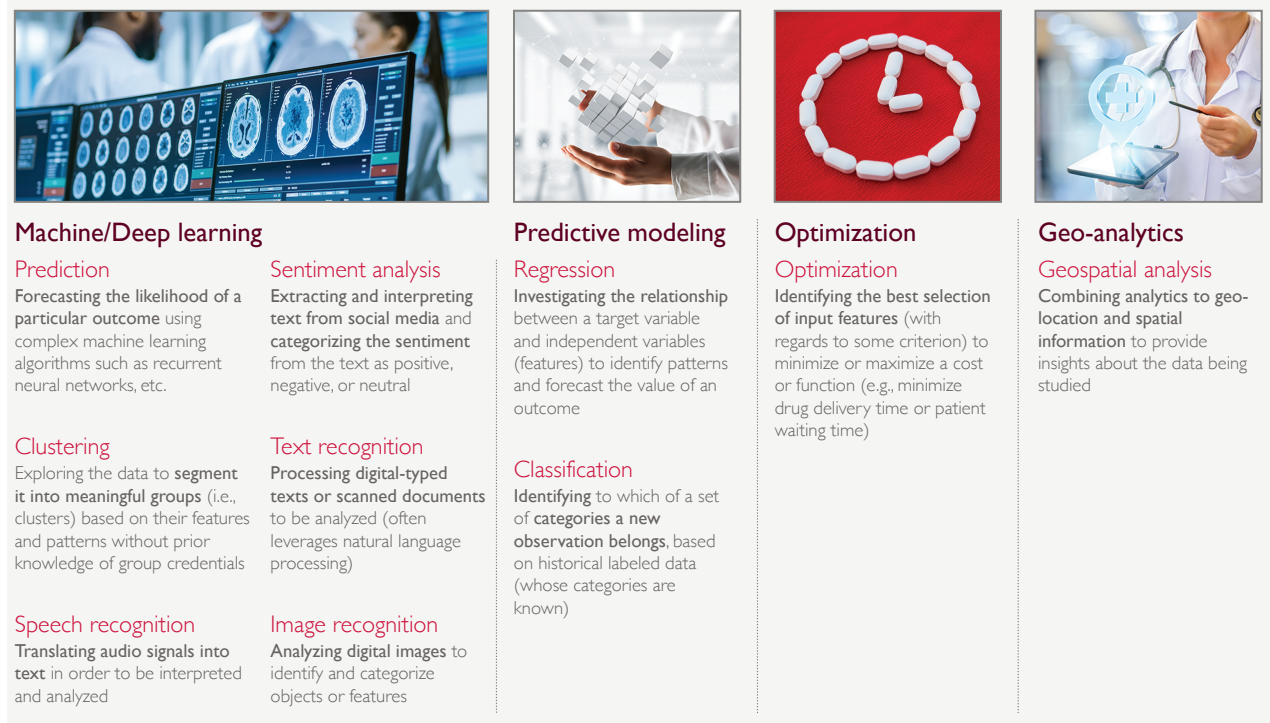
- **Clinical outcomes:** Number of PLHIV whose outcome is improved by the use case.
- **Cost efficiency:** Potential annual cost savings from the use case.

Feasibility was defined by the following factors:

- **Change management and sustainability:** Change to ways of working required and long-term funding and management sustainability.
- **Data condition:** Availability of, access to, and quality of required data.
- **Data science feasibility:** Possibility of the use case working successfully, based on previous, similar projects.
- **Tech/IT requirements:** Software and hardware requirements to implement and run the use case at scale.

Please see the appendix to this report for more detailed descriptions of impact and feasibility sub-criteria with allocated weightings based on relative importance.

Figure 4: Advanced analytics techniques considered in this analysis



The forty use cases in the long list were rated as low or high impact and feasibility, with a high-level rationale behind the choice, based on a brief description of each use case. Having key stakeholders make a binary choice ensured that the majority of use cases were not rated as medium impact or feasibility. Tradeoffs were made to ensure a balance of use cases across the patient journey and the overarching categories (data management, operations management, and staff). In addition, apart from the clearly high-impact and high-feasibility use cases, a mix of use cases with high impact and relatively low feasibility and those with low impact and relatively high feasibility was fostered. These “big bets” versus “quick wins” were important to help test a variety of different approaches in the more detailed assessment in the next section.

Please see the appendix to this report for a description of each use case in the long list.

5.3 Ten impact-generating use cases

Ten use cases emerged from this initial prioritization and are shown along the patient journey in Figure 5. Please keep in mind that the use case prioritization could very much differ depending on a country or organization’s specific priorities. The detail of these characterizations— for example, the data available or the specific end user— also varies by context. However, these use cases are

likely to resonate in some form for every context due to the almost universal pain points that they address in the HIV landscape and their potential to deliver significant impact on outcomes and cost efficiency.

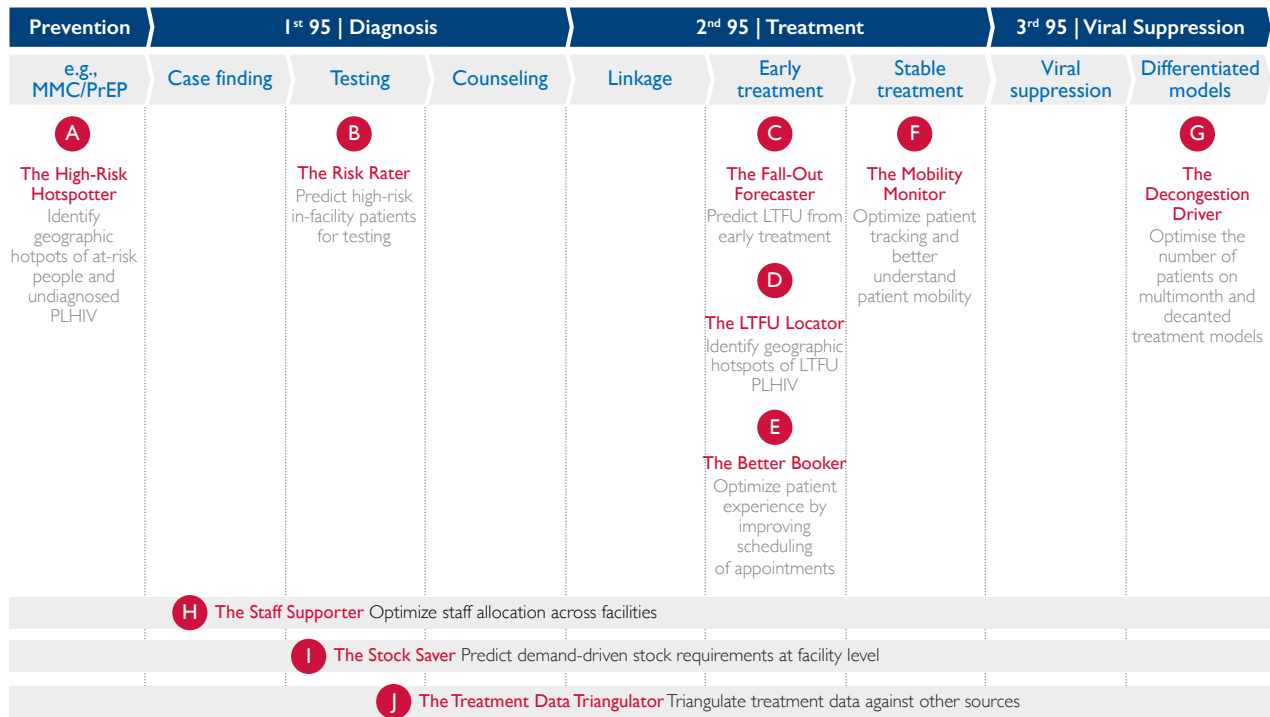
This section provides an overview of the top ten use cases with a description, a value statement (how the use case helps us get to 95-95-95), an illustration of how the use case would work in practice, and an estimate of the possible impact. We also show how these use cases could be applied to the pain points experienced by Ruth and Gift described in a previous section. Please note that the following descriptions are meant as an initial guide and would be refined with exact process and data requirements further on in the use case delivery model. Impact projections are particularly context dependent and depend on effectiveness of implementation and change management.

A. The High-Risk Hotspotter

Description: As a program manager, I want to better understand risk profiles for HIV incidence and incidence hotspot locations and have up-to-date alerts on area or population factors such as crime, power outages, and other events to help optimize community health worker (CHW) work.

Value statement: Identifying and characterizing hotspots enables incidence prediction and targeted prevention

Figure 5: Ten impact-generating use cases along the patient journey



initiatives in high-risk areas, which will also help to diagnose PLHIV in a shorter time period and more cost-effectively.

Impact: Potential incremental annual reduction in incidence of 3%-7%, increase in identifying/diagnosing new patients of 6%-11%, and cost savings of roughly 5% of relevant prevention and testing spend.

End-to-end process: How the High-Risk Hotspotter would operate in practice is shown in Figure 6.

B. The Risk Rater

Description: As a facility manager, I want to predict which people presenting in health care facilities have a high chance of being HIV positive, to ensure that they are approached for testing and that testing resources can be used more efficiently.

Value statement: In some LMICs, universal testing goals are being met in terms of numbers tested, but yield is insufficient; testing in-facility provides an immediate linkage opportunity for PLHIV who would otherwise remain untreated.

Impact: Potential increase in new diagnoses of 8%-13% from higher testing yield, increases in rates of patients returning to care of 3%-5% annually, and cost savings on testing spend in the range of 10%-15%.

End-to-end process: How the Risk Rater would operate in practice is shown in Figure 7.

C. The Fall-Out Forecaster

Description: As a case manager, I want to predict the likelihood of a patient dropping off or interrupting treatment within the first twelve months and, based on the patient's risk factors and individual behavior, optimize support through focused interventions.

Value statement: Keeping PLHIV on treatment throughout their first year is a key focus area to reach 95-95-95.

Impact: Potential reduction in early LTFU of 6%-10% and direct savings of 4%-5% on care and adherence support costs.

End-to-end process: How the Fall-Out Forecaster would operate in practice is shown in Figure 8.

D. The LTFU Locator

Description: As a program manager, I want to identify geographic hot spots of LTFU PLHIV to target priority areas for retention and back-to-care interventions.

Value statement: Being able to pinpoint factors that cause LTFU in areas or groups will enable proactive intervention to prevent it or to get PLHIV back to care.

Figure 6: High-Risk Hotspotter from input to usage

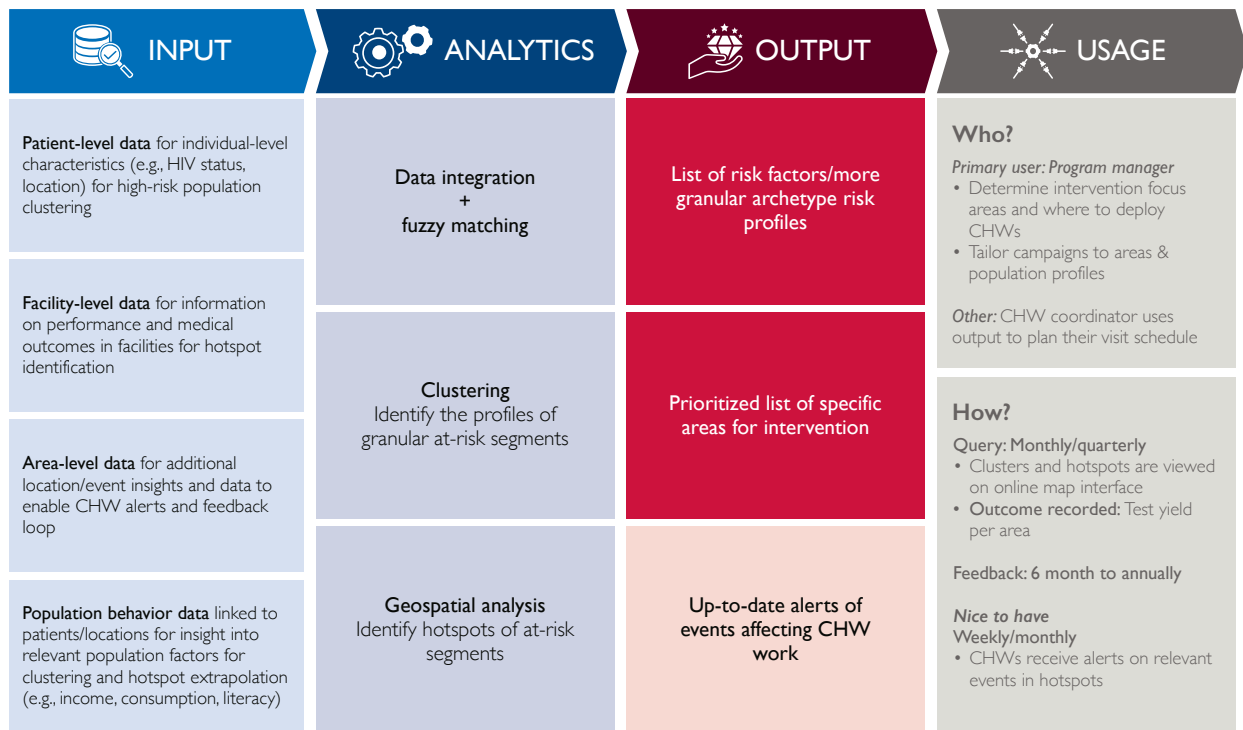


Figure 7: Risk Rater from input to usage

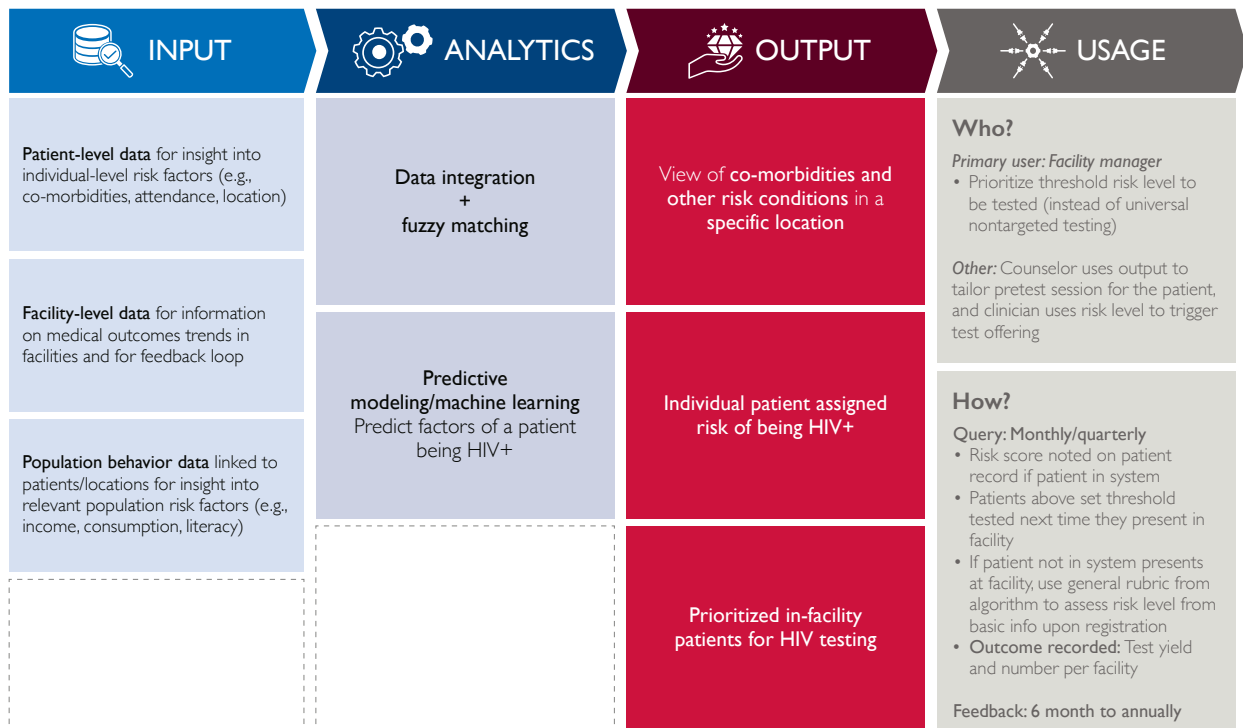
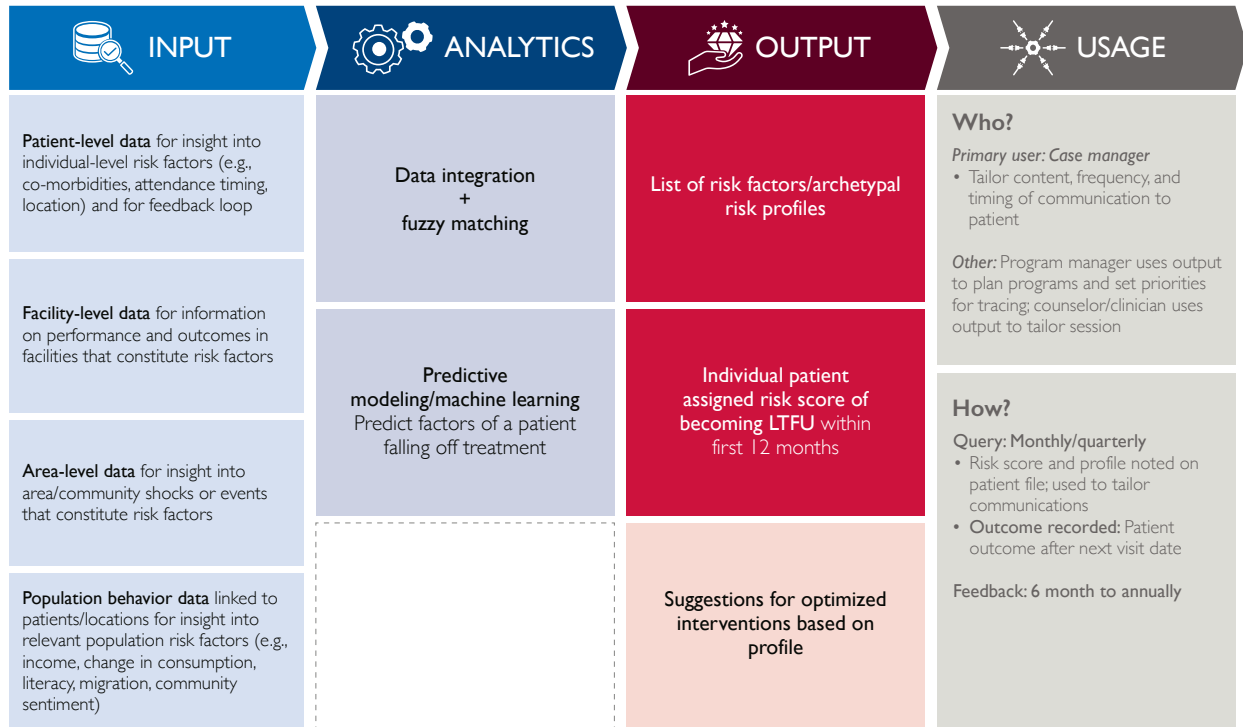


Figure 8: Fall-Out Forecaster from input to usage



A DAY IN THE LIFE: GIFT

Gift went to the hospital at the beginning of the year after he broke his arm playing soccer. Unbeknownst to him, he contracted HIV a year before from his long-term girlfriend, Nasneen. She slept with an older man during their relationship, which ended a few months ago when Gift found out about her affair.

When Gift visited the hospital, the doctor who fixed his arm suggested that he get an HIV test. The doctor noticed in Gift's medical record that he had been flagged as being medium-high risk of being HIV positive. The sub-district where he lived had a high incidence versus the rest of the country, and he had been treated for gonorrhea a few years ago at a different facility.

Gift was shocked to find out that he was HIV positive and decided not to tell his friends, family, or anyone at work. At his second appointment, his case manager suggested that he see a specialist counselor, who was also an HIV-positive young man, and asked if he'd like to communicate weekly with the counselor over WhatsApp. Unbeknownst to him, Gift had been flagged as medium-high risk for dropping off treatment, so special efforts were being made to ensure this didn't happen.

Impact: Potential reduction in LTFU of 3%-6% annually and potential direct savings of 3%-5% on care and adherence support costs.

End-to-end process: How the LTFU Locator would operate in practice is shown in Figure 9.

E. The Better Booker

Description: As a facility manager, I want to improve overall patient experience and minimize facility waiting time by enabling patients to select an optimal timeslot for an appointment from a range of options, based on facility capacity and the date of appointment.

Value statement: Long waiting time is a key patient satisfaction pain point. Addressing it could drive better clinical outcomes and retention across the care cascade and improve staff productivity.

Possible impact: Potential reduction in LTFU of 8%-12% per year, increase in PLHIV initiated on to treatment of 5%-8% a year and potential direct savings of 1%-4% on costs associated with tracking, case management, and workforce allocation.

End-to-end process: How the Better Booker would operate in practice is shown in Figure 10.

F. The Mobility Monitor

Description: As a case manager, I want to optimize tracking of patient mobility between facilities, districts, and provinces, to not miscategorize “silent transfers,” to provide uninterrupted support, and to enable efficient referral processes.

Value statement: Preventing LTFU is a key priority for 95-95-95; since mobility is often a driver of LTFU, supporting patients through mobility and also correctly recording when patients have moved rather than fallen off treatment can help us prevent LTFU and allocate resources for retention interventions more effectively.

Impact: Potential reduction in LTFU of 1%-3% per year from more accurate categorization of silent transfers, additional ~1% reduction in actual LTFU, and potential direct savings of 3%-5% on care and adherence support costs.

End-to-end process: How the Mobility Monitor would operate in practice is shown in Figure 11.

G. The Decongestion Driver

Description: As a clinician, I want to optimize the decision of which multimonth or differentiated treatment model a patient should be on, to accelerate use of less

Figure 9: LTFU Locator from input to usage

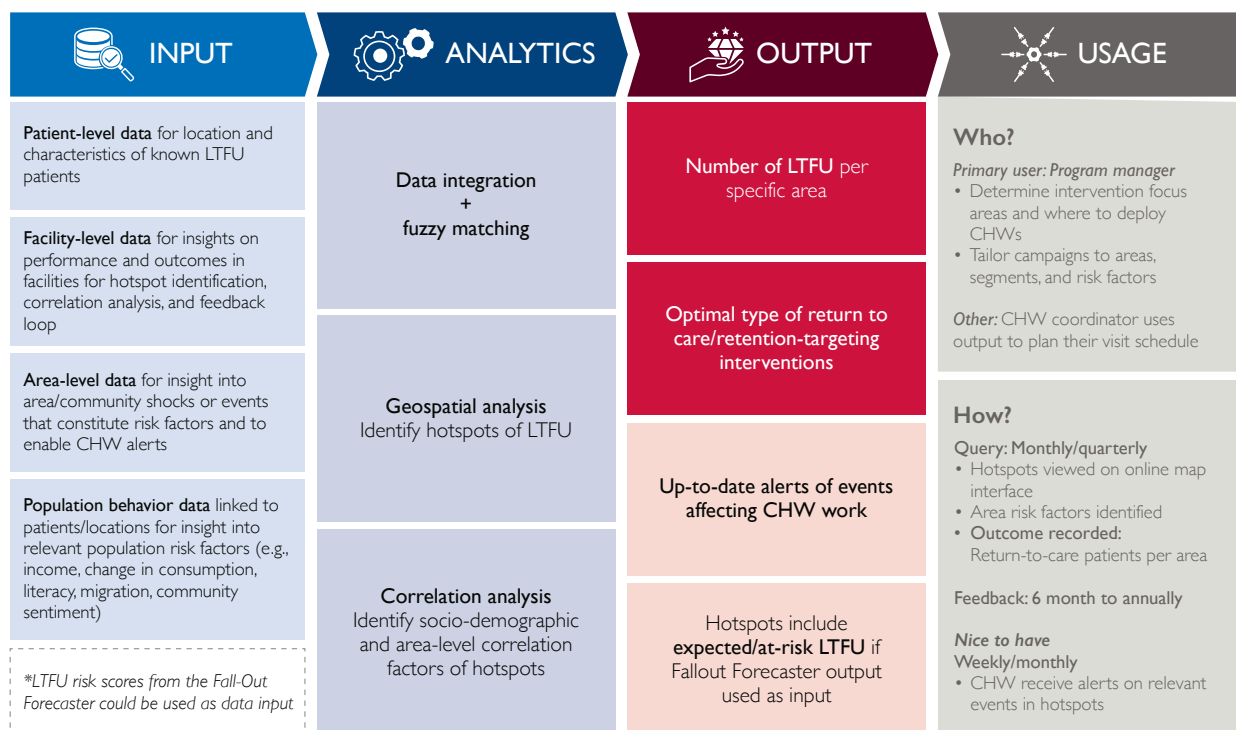
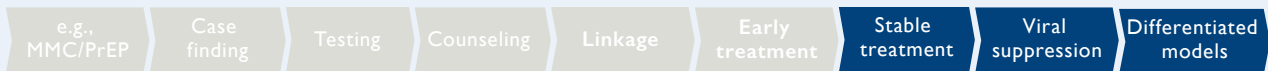
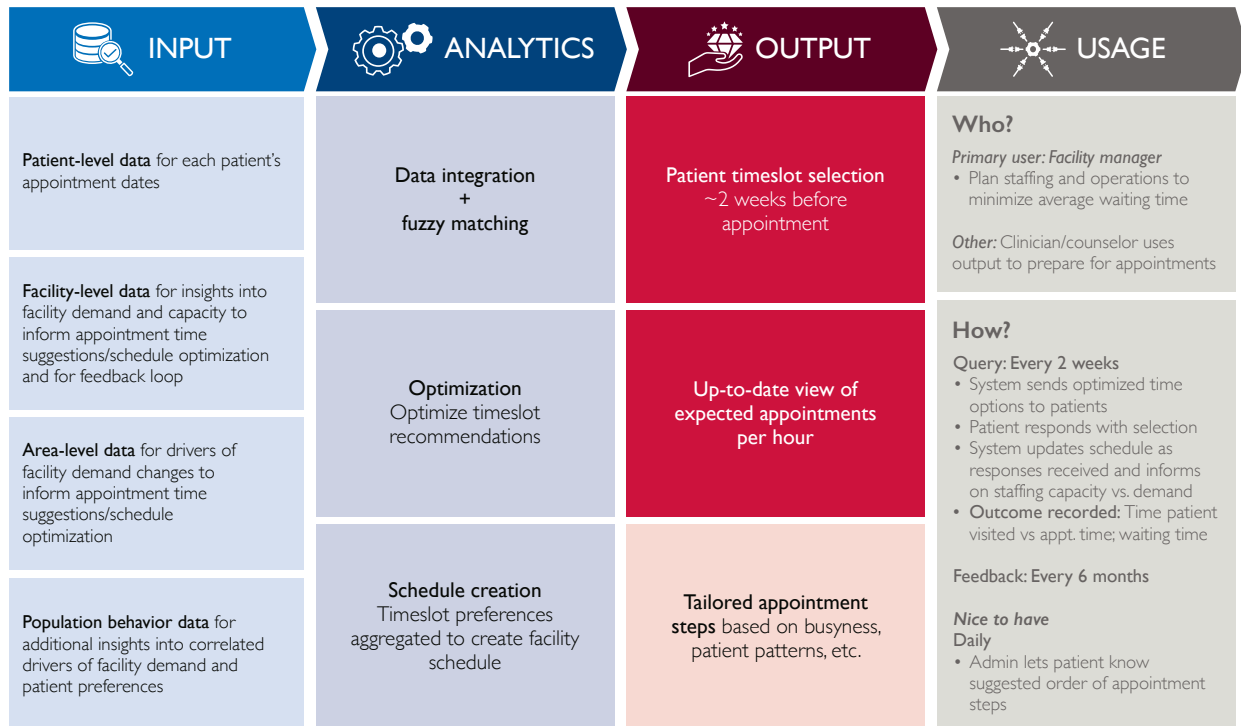


Figure 10: Better Booker from input to usage



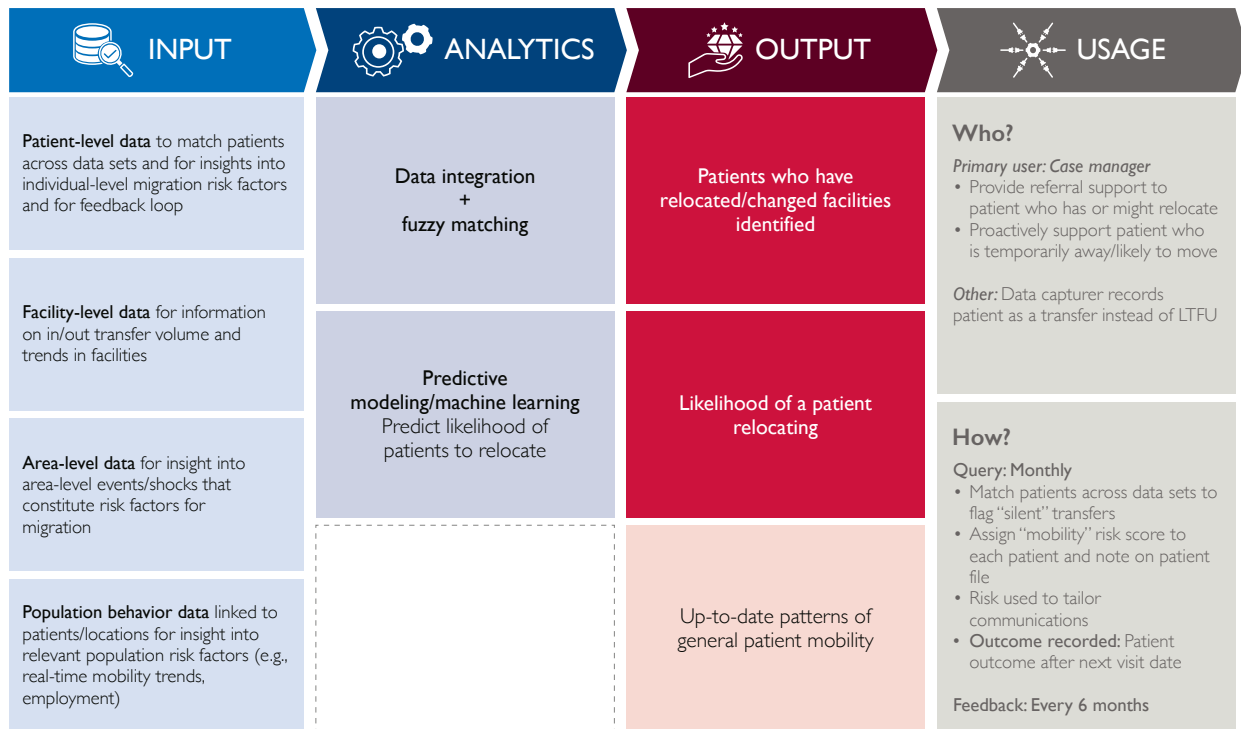
A DAY IN THE LIFE: GIFT

Two weeks before his appointment day, Gift receives an SMS from his clinic. He's relieved to see that the wording is generic and would not suggest to anyone reading it that he is HIV positive. The SMS requests him to choose a timeslot for his appointment, and says that if he arrives on time, he should not have to wait. Gift is doubtful—he's spent too much time in the queues with screaming children—but picks 1 p.m. Friday, a week from today, when he knows his manager will be at a lunch meeting. Two days before his appointment, he receives a reminder SMS. Gift arrives at the clinic five minutes before his scheduled time and is met by a new coordinator at the clinic, Ralph.

Ralph has been allocated to the clinic for a two-month period over December and January as the number of patients visiting is expected to increase significantly based on historical patterns of people making their way home for the holidays. Ralph marks Gift's timely attendance on the list and, after a few minutes, directs him to the doctor. Gift is pleasantly surprised by this process. Walking back to work, he reflects on the fact that he spent less than forty-five minutes in the clinic today, versus the usual minimum of two hours.

Ten months later and Gift is doing well. He has been visiting the same clinic since his diagnosis, partly because he can book his appointment timeslots there. This has made a big difference to him in managing his HIV, as he can plan appointments to suit his schedule—important since his promotion at work—and feels respected by the clinic staff who don't make him wait. Gift has an appointment with one of his favorite doctors, Dr. Ndashe. Today, Dr. Ndashe tells Gift that he would be a good candidate for external pick-up of his medicines. She explains that his viral load is suppressed, and based on his lack of other health issues, his young age, stable job, and a long list of other factors, he can collect his ART at the pharmacy in the mall next to his call center office. Gift is very happy with this outcome.

Figure 11: Mobility Monitor from input to usage



resource- intensive treatment models given adherence and cost benefits.

Value statement: Limited capacity in health care systems necessitates getting millions more patients onto less resource-intensive treatment models, and a clear case must be made to shift some patients' treatment plans with quality assurance.

Impact: Potential to transfer up to an additional 10% of treated patients onto less resource-intensive models of care, even if a country already has ~75% of patients on these models. Potential cost savings from in-facility cost avoidance are significant, ranging from 50%-80% reduced treatment cost per patient that is moved to a differentiated model depending on the number of facility visits still required.

End-to-end process: How the Decongestion Driver would operate in practice is shown in Figure 12.

H. The Staff Supporter

Description: As a human resources manager, I want to optimize staff allocation across facilities based on capacity requirements, predicted demand variability, and patient risk profiles.

Value statement: Better matching staff capacity and

capabilities to patient needs could have a significant impact on quality of care and cost reduction in the system from overtime and inefficiency.

Impact: Potential increase in number of PLHIV initiating onto ART of 1%-2% per year, potential 2%-4% reduction in LTFU, and potential direct savings of 4%-6% on human resources and adherence support costs.

End-to-end process: How the Staff Supporter would operate in practice is shown in Figure 13.

I. The Stock Saver

Description: As a supply chain manager, I want to predict ARV stock requirements per facility based on actual and forecasted patient treatment data and population trends.

Value statement: Stockouts and overstock cause significant direct and indirect costs. Stockouts are a driver of patient dissatisfaction and LTFU.

Impact: Potential reduction in number of LTFU of 1%-3% per year and potential direct savings of 2%-4% in ART procurement spend.

End-to-end process: How the Stock Saver would operate in practice is shown in Figure 14.

Figure 12: Decongestion Driver from input to usage

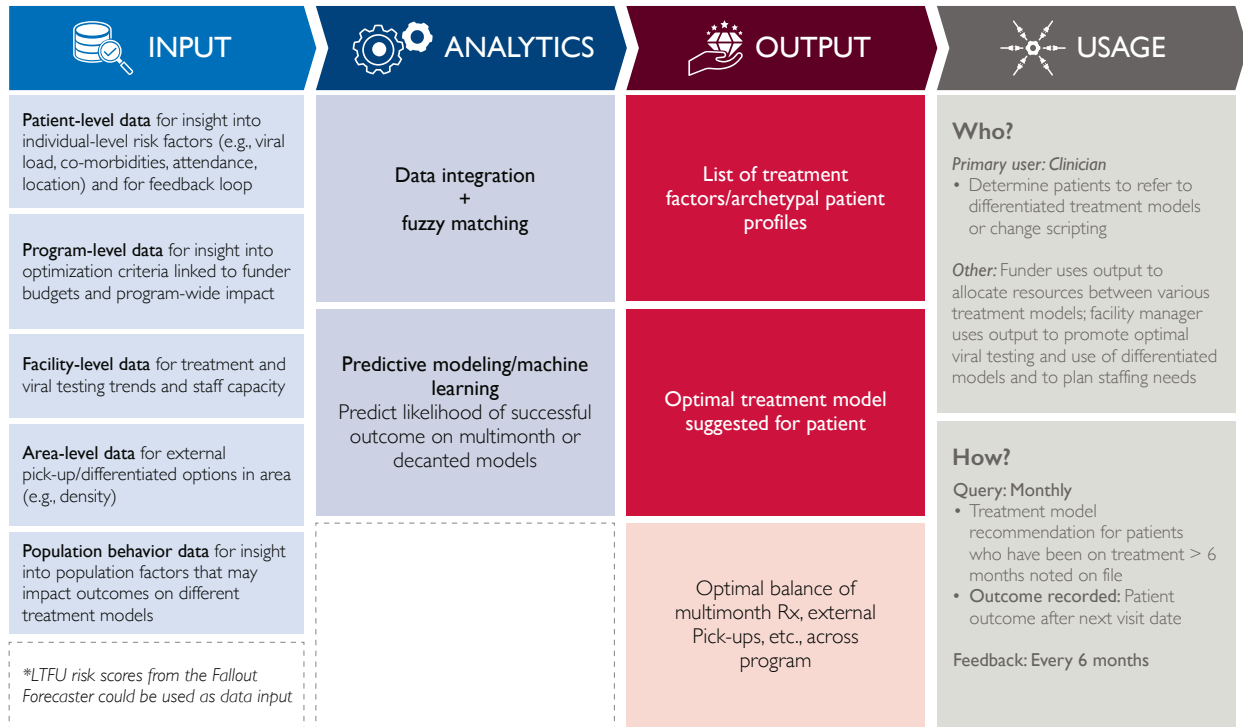
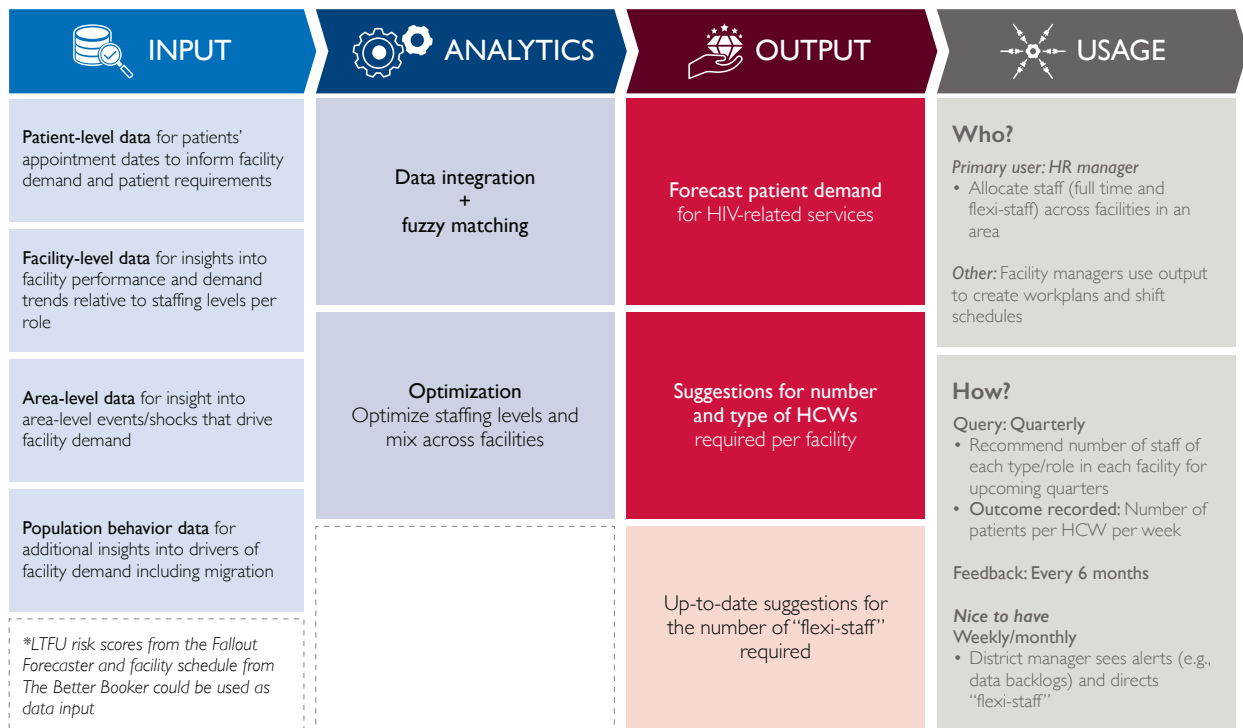


Figure 13: Staff Supporter from input to usage





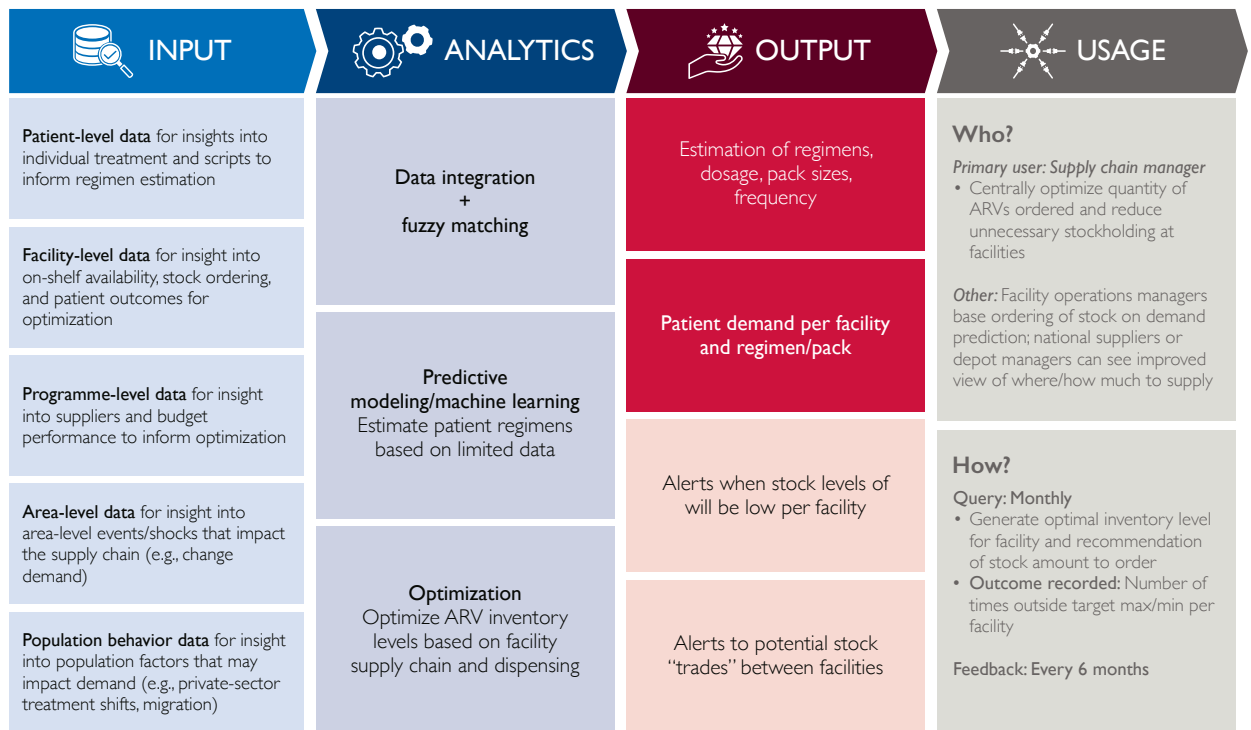
A DAY IN THE LIFE: RUTH

Ruth's patient, Ahmad, visited a different area of the country a month ago to care for his sick mother, but has not returned. Unable to get hold of him, Ruth logs a request to her district manager to check the Mobility Monitor for Ahmad's whereabouts. A few days later, Ruth received a call.

Ahmad has been found in the patient-level treatment data of a small rural clinic, close to where his mother lives. Ruth is relieved that Ahmad is still on treatment and marks a transfer in the system.

Ruth has some extra time on Friday mornings, and Lulu, the manager of one of the big retention programs in her district, has asked her to help out with a new campaign the program is running to get LTFU PLHIV back to care. A suburb in Ruth's district has been identified as a hotspot of elderly people who have dropped off treatment due to a nearby clinic closing. Lulu has designed a targeted program to support these people in accessing their medicine through a new adherence club. Ruth is going to help ensure they receive the support they need to get back onto treatment.

Figure 14: Stock Saver from input to usage



J. The Treatment Data Triangulator

Description: As a program manager, I want to triangulate treatment data with other relevant data sources to have a better understanding of delays in data capturing and other data quality challenges that skew patient treatment numbers.

Value statement: Trying to understand data issues and fix the quality of data takes significant time and effort from implementing partners.

Impact: Limited direct impact on clinical outcomes but potential direct savings of 8%-12% on human resources spend related to data capturing.

End-to-end process: How the Treatment Data Triangulator would operate in practice is shown in Figure 15.

5.4 Key learnings

Describing the use case, refining the value statement, and creating an end-to-end process flow will enable a more in-depth impact and feasibility assessment for the ten priority use cases. One overarching learning is that more of the ten priority advanced analytics use cases are classified as “big bets”—higher impact, but relatively lower feasibility—than “quick wins”—higher relative feasibility,

but lower impact. The “big bets” use cases pose exciting opportunities to potentially reach hundreds of thousands of PLHIV or to save significant cost in the HIV landscape, but require cooperative effort from stakeholders across the HIV value chain to address their feasibility challenges and make them a widespread reality. Although the “quick wins” might seem more tempting in the short term, “big bet” use cases can often tackle challenges that may be intractable without a digital element and therefore should be a focus. However, “quick win” use cases offer a valuable opportunity to build momentum with some early wins if your organization or country is embarking on a broader digital transformation.

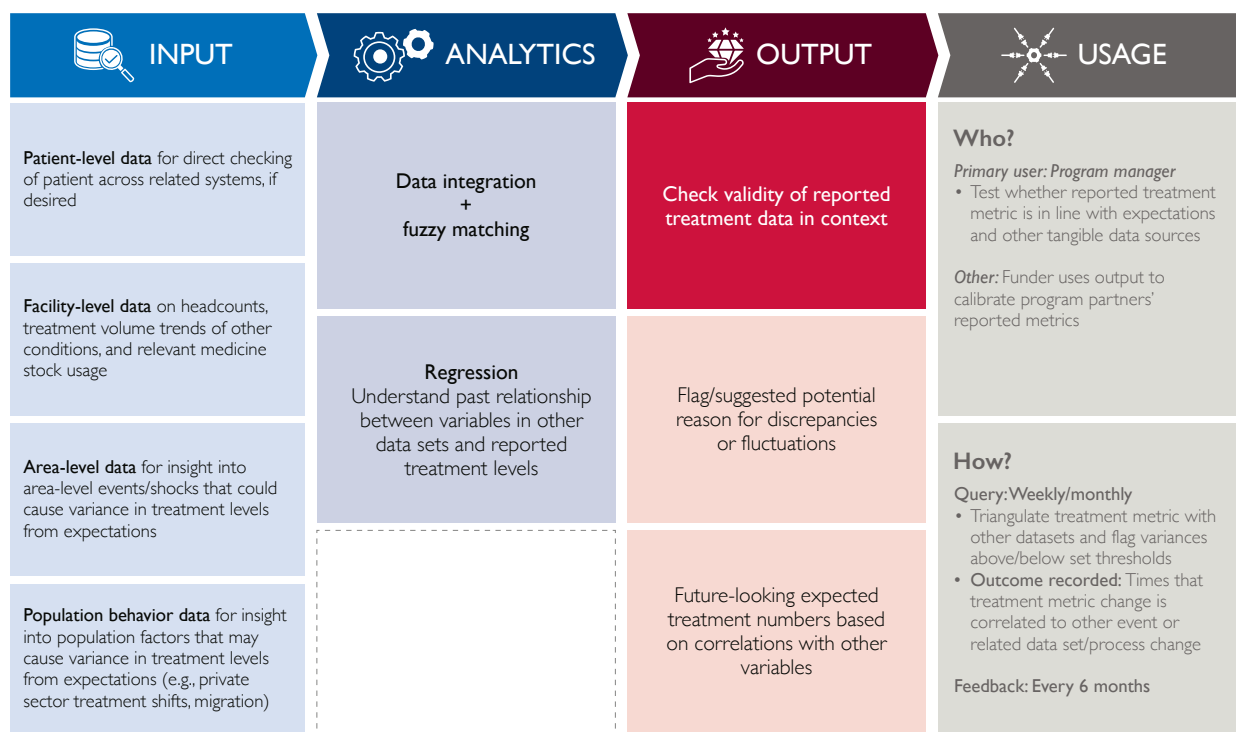
5.4.1 Factors affecting impact

There is no doubt that advanced analytics can have a major effect on the HIV landscape. In fact, we came across very few pain points that could not be addressed in some way through a use case. The key learnings below reflect the potential impact of the use cases, as well as process learnings that helped us design the use cases in the most effective way possible.

Use contextual priorities to direct use case focus

Although the ten priority use cases should be valuable in most countries, the extent of the impact will depend

Figure 15: Treatment Data Triangulator from input to usage



on a country's progress toward 95-95-95 and the constraints it faces in achieving those targets. A key consideration is whether the focus looking toward 2030 is predominately on initiation of PLHIV onto treatment, or retention of these patients on treatment, to achieve viral suppression.

In addition, government and funder priorities and existing projects should be considered to ensure buy-in and leverage impact. Use cases should be developed around context-specific pain points and priorities. For example, the Better Booker could play an important role in supporting initiatives to improve facility operations. The Better Booker would provide patients with an optimized appointment timeslot, rather than just an appointment day, which has been shown to significantly reduce waiting time and increase patient satisfaction. As the transition to once-a-day, fixed-dose combination TLD ramps up in many countries,² advanced analytics use cases could be employed to nudge clinicians to transition patients earlier rather than later, and help supply chain managers balance the changing stocks and flows of ART regimens.

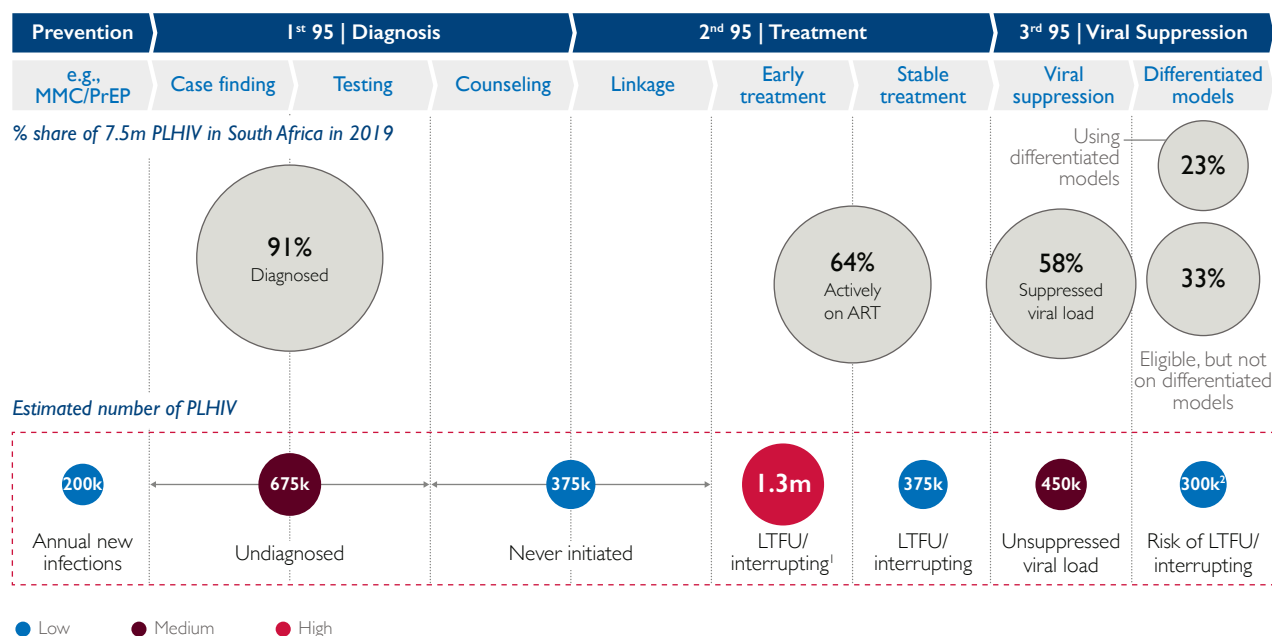
Quantify the number of patients affected at various points of the journey with comparable metrics

The impact of use cases can be evaluated by looking at a country's treatment cascade. A high-level analysis of size of patient pools along the journey helps to identify where the most impact from use cases would be realized. As an

example from the country with the largest HIV burden globally, South Africa, Figure 16 provides an illustrative overview of the number of PLHIV in each step of the patient journey. Although overly simplified, this graphic can help us understand the potential impact of a use case that targets a specific step. The key comes in translating traditional treatment cascade values (shown in gray in Figure 16) into the number of patients that are falling out of the cascade at different points (shown in yellow/orange/red in Figure 16). This allows you to then estimate in a comparable way how many patients' outcomes could be improved if a use case is implemented that addresses a specific point in the cascade/journey.

In this South African example, two hundred thousand new HIV infections in South Africa occurred in 2019, representing 3% of total PLHIV in the country. Of total PLHIV, 9% (675k) were undiagnosed, suggesting work remains to be done in increasing testing rates and, more importantly, testing yield. However, arguably the most problematic PLHIV pool is the 17% (1.3M) who interrupt or are LTFU from early treatment (first twelve months). Retaining PLHIV on treatment throughout their journey is important but especially difficult in the early phase of treatment. Impact in this critical step of the patient journey could be realized through a number of the top use cases we highlighted: the Fall-Out Forecaster, the LTFU Locator, the Better Booker, the Mobility Monitor, the Staff Supporter, and the Stock Saver. Creating a

Figure 16: PLHIV pools along the patient journey



¹ 80% of LTFU/interrupted ART allocated to early treatment.

² Proxy for potential impact of getting more patients into differentiated models.

comparable metric such as number of patients lost to treatment (or at risk of falling off if not for interventions, such as in the example of patients eligible but not on differentiated treatment models) can allow systematic prioritization across use case options.

Seek out (or produce) research quantifying the underlying drivers of issues

The size of the patient pool affected by a use case is only the first step to estimating overall impact. To truly estimate the impact of a given use case requires deep understanding of what is causing the issue for patients (e.g., why did they fall off treatment?) and then linking the use case features to which of these underlying issues are improved. Unfortunately, although general factors that cause LTFU, lack of initiation, or undiagnosis are fairly well understood at a qualitative level, the extent to which each factor quantifiably contributes—notably for specific groups such as men or adolescent girls and young women (AGYW)—is usually unclear. When assessing the impact of use cases that addressed LTFU, questions such as the following were relevant: In what share of LTFU does in-facility waiting time play a role? How many PLHIV have interrupted treatment due to stockouts as the major driver? Estimating impact is more powerful when this knowledge can be quantified. The research community can accelerate improvements in our understanding of behavior drivers through quantitative research using techniques such as discrete choice experiments³ rather than stopping at qualitative insights only.

Beyond number of patients affected, consider likelihood of impact and potential cost savings

The impact of each of these use cases depends on the probability that action is taken based on the use case insight (e.g., launching a retention intervention based on a patient being flagged as high risk for LTFU). Impact further depends on the intervention's likelihood of success. For example, how likely is it that the patient will respond to the intervention and remain on treatment? Historically, certain types of interventions have had different success rates, and assumptions from relevant research can be used to estimate this.

Improving PLHIV outcomes is the main goal, but doing it cost-effectively is becoming increasingly important as the global health community faces reductions in overall funding. Each use case was found to have some expected efficiency gains from, for example, more targeted deployment of testing kits, less required facility capacity from reducing patient in-facility visits through differentiated models, better deployment of facility staff roles to meet demand, or reduction of manual tracking and tracing costs. These gains can be estimated from prior applications of similar analytics use cases whether

in similar public health environments or from private-sector examples. Additionally, system-wide indirect cost reductions can also be realized, such as decreasing the cost of treating HIV-related complications and opportunistic infections arising from nonadherence to ART.

A use case's impact is dependent on it being feasible. If a use case is particularly difficult to implement, its impact may be reduced or at least delayed. The next section explores common feasibility challenges across the ten priority use cases.

5.4.2 Feasibility

Specific feasibility challenges encountered will depend on the context and objectives of your use case, as well as the specifics of the use case itself. However, across our analysis of the ten priority use cases, common feasibility challenges arose. These challenges and how to mitigate them are described in the following paragraphs.

Concentrate end users, ensure comfort with digital, and align incentives

Convincing people to change their entrenched ways of working poses a significant challenge to any use case. However, if the use case focus is on strategy or planning rather than operations, it is likely to have fewer, more concentrated users: one program manager, for example, rather than multiple clinicians. If the end user is more used to working with digital (e.g., if the end user is a monitoring and evaluation manager), there might be better take-up. Ensuring that incentives are aligned to encourage the end user to use the tool is another way to reduce change management issues.

Centralize the use case, and avoid running it in real time

A useful learning is the possible feasibility boost available from centralizing use cases and running them as frequently as necessary but not in real time. Centralizing a use case refers to where and how it is used; for example, the Risk Rater is best run by a central facility manager, although patients' HIV risk scores would then be included in an existing database or file accessible by clinicians. This requires less widespread change to ways of working as compared to having each clinician accessing and running the use case. Centralizing a use case also leads to less-intensive IT/technology requirements, notably less end-user hardware and widespread in-facility Internet connectivity. Running a use case in real time to obtain a result for an individual instance requires significantly more processing power than running a use case at a set frequency to obtain results for a group that are recorded for later use.

Avoid including input data from on-the-ground surveys

Multiple use cases would benefit from collecting information from patients, to enrich the view of their risk, or from facilities, to better understand current processes and outcomes. For example, up-to-date information about a patient's sexual history and behavior would be very valuable for a use case predicting a person's risk of having HIV. However, the time, resource, and operational requirements of collecting even short in-facility surveys can be debilitating for feasibility and should be avoided wherever possible. Implementing a use case already requires changes to entrenched ways of working; if an additional, time-consuming step is added, the chance that the use case would be used regularly decreases significantly.

Partner with the owners of patient-level data sets

Health systems' strengthening is the overall goal in implementing advanced analytics use cases. Integral to this is partnering with data owners in a way that ensures their priorities are addressed and that the access needs are clear. This is particularly important considering that patient-level data sets, individual records of tests, treatment history, and outcomes are integral for the impact of almost every use case we considered. Using patient-level data introduces complexity. Globally, there is increasing sensitivity and detailed requirements for

access, use, and storage of such data. Patient-level data is often owned by government at different levels, and accessing the data requires numerous permissions. Even if de-identified patient data can be used, restrictions remain. A major consideration is how frequently, and in what way, the data set needs to be accessed. If a data set needs to be accessed regularly for the ongoing running of the use case, or if read and write access is required, this can introduce additional complications. These complications can be mitigated by building use cases in direct partnership with the data owner(s).

Determine data requirements early on

Some use cases would benefit from rich process data, such as records of appointment times, duration, and attendance, which isn't necessarily available. In other cases, outcomes data to test the correlation of input factors against is lacking. For example, a use case that predicts a patient's optimal treatment model based on the patient's individual profile and risk factors would initially require data linking patients' profiles and treatment models to outcomes such as retention or dropout and, even better, patient satisfaction metrics. Where this data has not yet been generated or collected, the implementation of the use case may be delayed until a sufficient number of data points is available. This requires engagement from varied stakeholders in the framing stages of the use cases, so should be determined early on.

Footnotes

1. Ratings have been simplified for the purposes of this report.

2. TLD stands for tenofovir disoproxil fumarate, lamivudine, and dolutegravir.

3. Zanolini, Arianna, et al., "Understanding preferences for HIV care and treatment in Zambia: Evidence from a discrete choice experiment among patients who have been lost to follow-up," *PLOS Medicine* 15 (2018):e1002636. [10.1371/journal.pmed.1002636](https://doi.org/10.1371/journal.pmed.1002636).

6. Data sources, challenges, and opportunities

6.1 Patient-level data in LMIC countries

Patient-level treatment data is integral for almost every use case we considered, and the success of a use case depends on the quality of this data.¹ Challenges in data collection, storage, and management are important to address, as they can result in significant skews in use case analyses. Operational and programmatic decisions, such as capacity requirements for tracking and tracing, depend on robust patient-level treatment data as well. However, in many resource-constrained countries with a high HIV/AIDS burden, issues with the quality of this data arise along the following steps of the data journey:

- **Collection:** Recording of raw data, for example, in patient files or online/offline systems.
- **Storage and management:** Capturing raw data into databases and long-term management of data.
- **Aggregation and processing:** Integration of datasets, indicator calculations, and quality validation of data.
- **Reporting and interpretation:** Evaluation of calculated outputs and interpretation and use of results.

The data systems landscape of LMICs reflects challenges that are common in transitioning away from legacy and paper-based systems:

- 1). A hybrid of paper-based and electronic systems is used at most facilities.
- 2). The public health care database ecosystem is highly fragmented.
- 3). Often, a large amount of the facility data team capacity is taken up by highly laborious tasks, such as loading patient information captured on paper-based systems into electronic systems.

- 4). Significant efforts are spent on validating and cleaning data for reporting requirements.
- 5). Difficulties in performing deeper analyses occur as a result of using legacy systems not designed for their current use.

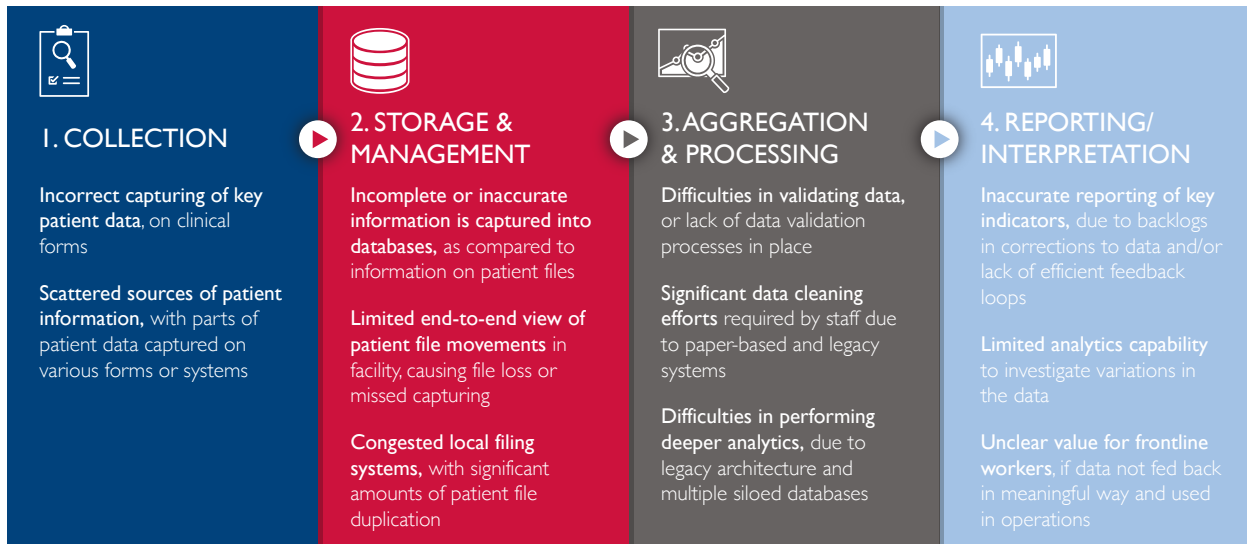
Some of the major challenges LMICs often face in producing reliable patient-level data are summarized in Figure 17. When considering issues with data governance, it is important to take a user-centric view of the root causes behind the issues you see. Although the symptoms can present differently, the root causes can generally be split into two categories:

- **Behavioral-related causes:** This refers to the perceived value of performing tasks and the associated approach to performing them. For example, data captureurs may place more value on capturing all patients' files into the system than on accurately capturing the data, or clinicians may not realize, when they record an important element in one register but not the other, that the reporting counts are actually based on the register they skipped.
- **Capacity-related causes:** This refers to people's ability and capacity to perform data-related tasks. Examples include giving clinicians too many different registers to fill but also limiting the time they have to see patients, or having data teams who are not resourced adequately with clear tools or guidance to perform effective reviews of data quality at their facilities. This can be exacerbated by misaligned or complex indicator definitions and data exports, which can lead to a lack of understanding and lack of use by staff on the ground.

Quality issues in patient-level data can be addressed by considering four key questions across the data journey:

- 1). What current processes need to be adjusted or new processes implemented?

Figure 17: Examples of patient-level treatment data issues in each phase of the data journey



- 2). How should I change the way I look at my data/ use my data?
- 3). How can I leverage my existing workforce (people) more efficiently?
- 4). Do I have the right tools in place to support my processes?

Figure 18 is a generalizable map of key elements that should be in place across the data journey to mitigate the issues and root causes as much as possible given constraints.

For each key question, common best practices have been identified to significantly improve data quality:

Processes:

- Limiting the number of forms or registers that a clinician must fill per consultation to two to three, where possible.
- Creating an integrated system (even if paper based) linking patient attendance registers to a patient file tracking system that gives an end-to-end view of all patient file movements across the facility for easier validation and tracking.

Data:

- Performing audits or checklists on paper-based systems before loading the data into databases, such as making sure that the next appointment date for patients matches the number of months prescribed or an exception is signed off.

- Performing regular data quality checks on completeness, accuracy, business logic, and correct calculation.

People:

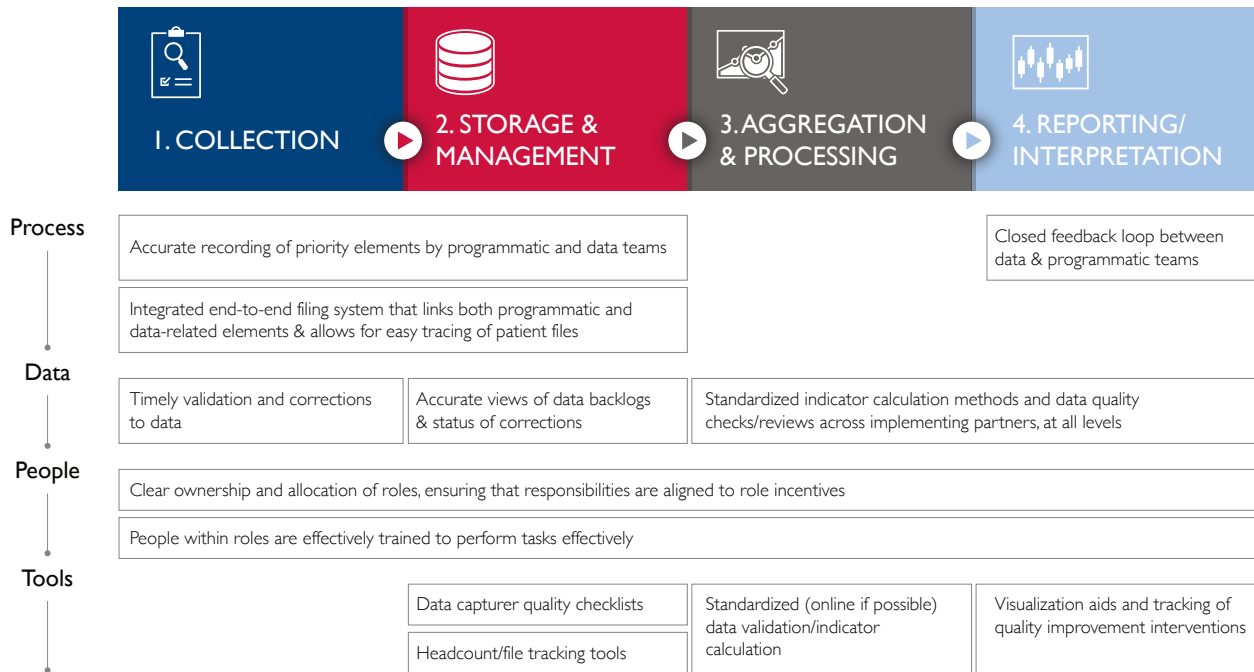
- Allocating clear roles and ownership of certain data elements to people whose role incentives align to that responsibility (for example, ensuring that the file clerk manages and owns a file tracking tool, regardless of the file clerk’s seniority).
- Bringing programmatic teams closer to data elements and driving the adoption of cross-functional teams, such as ensuring rapid and robust feedback loops from track and trace teams into facility data capturing teams.

Tools:

- Developing a data quality metric tool to use at the facility level to enable easier ownership and management of data quality and establish agreed-upon quality performance indicators that can be used for accountability.
- Creating a simple tool to support visualizing key data elements in an effective way to aid programmatic discussions on data, including interpretation of data quality issues and triangulation with multiple sources.

In addition to these elements, it is critical to ensure that data quality improvement processes are embedded in the national government’s objectives and systems, which helps in ensuring sustainable adoption.

Figure 18: Key elements to get right to improve data governance



6.2 External data considerations and sources

As seen in the end-to-end process diagrams for the top use cases in Section 5.3, the types of data used for advanced analytics use cases in health fall generally into one of four categories:

- 1). Individual data – for example, patient-level (or other individual-level) information, such as treatment records, laboratory results, or demographic information
- 2). Facility data – aggregated information at the level of facility, either medical or nonmedical facilities (e.g., supply chain depots), usually covering specific program metrics such as number of patients treated for a disease or number of staff at a facility
- 3). Area/population data – data providing insight into population characteristics and behaviors such as socio-demographic data, social media, retail, and media consumption, etc., or information on environmental factors such as weather, traffic, and crime
- 4). Program data – aggregated information describing programs, usually budget or expenditure related, and often split at the level of intervention

Most health programs have good collection, understanding, and usage of data categories 1, 2, and 4 (notwithstanding the challenges with individual patient-level data described in Section 6.1) but the biggest blind spot is often around area/population data. This gap is where external data sources (e.g., commercialized, academic, or private-sector data sets) can be incredibly powerful to shed light on relevant characteristics and behaviors of the communities that are served in and around health facilities. Direct linkage of individuals from health systems data to external data sets is not commonly possible due to regulatory requirements and data management practices of potential data partners, but much impact can still be gained from aggregated, de-identified data providing insight into community-level behavior. This level of data can be linked back to patients based on location information, enabling applications that can predict behavior due to location-specific factors such as socio-demographic profiles or providing population trends that shed light onto likely actions of people within a community.

A key finding from our landscaping of the external data ecosystem is that rich sources of external data pertinent to the top use cases are available and can be accessed in a variety of ways. Multiple data-driven businesses provide data sets and analytics platforms for purchase or subscription, and other organizations, such as global health and research institutions, make information publicly available for free or through low-barrier partnerships.

Many private-sector organizations are searching for ways to use their data for social good and have demonstrated a willingness to provide data access in the pursuit of mutually beneficial outcomes: improving public health while providing insight into population trends and behaviors that affect their own businesses. For external data partnerships to be successful, the value proposition for both entities needs to be clearly articulated and considered in the establishment of data-sharing agreements. Some examples of the value considerations that have been important to data providers include (i) a two-way sharing of data, including potential access to the pool of integrated data from other partnerships; (ii) sharing of data skills and capabilities; and (iii) demonstrated/quantified benefit that the output of a use case will have on their target population/key customer groups. Alignment on such details is important to ensure the interest of all parties involved are protected, and to prevent contractual or collaborative challenges that may detract from the ultimate goal of societal impact.

Aggregated data is easier to share than individual data but still requires careful consideration of data privacy in accordance with local regulations such as the General Data Protection Regulation in the EU or the Protection of Personal Information Act in South Africa. These regulations place very stringent guidelines on how individual-level data should be treated, and this needs to be respected and navigated with care in all partnership discussions. Often data-sharing consent must be given to the data aggregator by the original data owners/individuals (when data is sourced from other organizations), even when dealing with aggregate-level data sets. In any case, robust data governance measures must be put in

place before any data (individual level or aggregated) can be sourced and brought into the organization. Also, often external data requires establishing partnerships with organizations whose primary business is not data collection or analysis, and in those cases where partners have limited experience in sharing data or using their data assets for more than internal purposes, it is even more critical to ensure careful management of data privacy and governance.

In this study, we identified sixteen categories of external data with the potential to provide valuable input into our use cases and that have been observed to have growing applications for global health challenges. These data categories fall across health and non-health population data as well as area information. Figure 19 provides an overview of the categories and how they may be applied to use cases. Details on the specific categories are in Section E of the appendix.

Given the range of data set options that are available in multiple categories, it is important to identify the best data set required for each use case, considering the granularity of data, breadth of variables, data history, and frequency of updates required. Further, consider the ease and cost of access of each available data set. The key point of departure for decisions related to external data should always be first identifying the data-driven applications that are most beneficial for program outcomes, as we have done above for the ten priority use cases, then quantifying the impact of the use case and assessing that against the investment required to either purchase, subscribe, or partner for data access.

Footnote

1. For the purposes of this use case, treatment data is defined as the number of adults and children currently receiving ART.

Figure 19: Multiple categories of external data can provide additional insights for some of the use cases

	POPULATION DATA: NON-HEALTH RELATED			POPULATION DATA: HEALTH RELATED		AREA-LEVEL DATA		
	Socio-demographic	Consumer retail	Telecoms	Private-sector records	Private pharmacies	Traffic	Crime	Weather/power outages
The High-Risk Hotspotter	Critical insight into profiles and potential predictive factors of at-risk populations	Insight into risky behavior or challenges that may increase risk of infection and/or not getting tested	Corridors of movement of high-risk populations from community to facility			Busy areas for marketing and area disturbance alerts that impact CHWs ¹	Insights for tailoring messages in hotspots with high HIV-related ² crime	Alerts of area disturbances that could impact CHW ¹ outreach
The Risk Rater				Insights into comorbidities of PLHIV	Insights into individual or population medication factors predictive of being HIV+ or nonadherent			
The Fall-Out Forecaster	Potential individual or community/population-based predictive factors of falling off treatment	Population financial and consumption trends/shocks that are predictive of challenges to stay on treatment	Dwell time of patients at a facility and general busyness of facility	Insights into drivers of adherence on DMP ³				
The LTFU Locator				Insights into drivers of adherence and locations where public/private sector crossover is occurring		Busy areas for marketing and area disturbance alerts that impact CHWs ¹	Insight into potential challenges facing CHW in hotspots	Alerts of area disturbances that could impact CHW ¹ outreach
The Better Booker	Potential factors for tailoring appointment times					Facility demand at different times of day		Potential risk factors for missed appointments to optimize around
The Mobility Monitor	Insight into social factors that cause population mobility	Corridors/times of movement within a network of retailers	More precise tracking and prediction of patient population mobility		Corridors/times of movement within a network of pharmacies	Insights into changes in traffic that may indicate significant mobility		
The Decongestion Driver	Potential predictive factors of outcomes on different treatment models				Insights into chronic patient adherence on varying scripts			
The Staff Supporter			Patient dwell time and mobility trends			Facility demand at different times of day	Insight into drivers of facility demand and staff performance/needs	Insight into drivers of staffing trends (e.g. absenteeism)
The Stock Saver	Insight into population factors correlating with treatment regimen			Swings in area private-sector supply/demand, particularly in relation to public/private-sector crossover				
The Treatment Data Triangulator			Population mobility in comparison to reported treatment levels	Changes in area private-sector treatment rates				Potential reasons for discrepancies in captured/ reported treatment stats

¹ Community health workers.

² Related crimes include sexual violence and domestic abuse that are related to an elevated risk of HIV.

³ Disease management program.

7. Next steps: Key considerations for developing a use case

Once you have a prioritized list of top use cases for your context, and understand how sources of external data could provide richness, what comes next? We can think of the prioritized use cases as a “portfolio,” similar in some ways to an investment portfolio. Your organization’s use case portfolio should be well-balanced, well-timed, and well-managed. There might be some higher-risk use cases (big bets) and some safer options (quick wins), but each use case should meaningfully contribute to achieving your core objectives. The big bets, use cases with higher impact but lower feasibility, will have a significant payoff if they work. The quick wins, use cases with higher feasibility but lower impact, enable an organization to build digital momentum and potentially realize cost savings to fund the delivery of other use cases.

However, it is not possible, nor desirable, for the vast majority of organizations to consider developing many use cases in a couple of years. Bringing an advanced analytics use case to life is an intensive process that can take longer than a year from the ideation phase to having a fully deployed and scaled-up product. However, using an agile, sprint-based approach to development enables impact to be seen from an early stage in testing the product, and allows for course corrections along the way. Figure 20 shows the five phases of the use case delivery model, with a high-level timeline for each phase. This timeline does not consider possible delays in procuring a digital partner for implementation and getting buy-in from key stakeholders. The latter is particularly relevant in global health, due to the high number and variety of stakeholders and the level of engagement required.

Several key considerations are involved when deciding how to move forward. These will be context dependent, but could include how to collaborate with and enable partners in the health system to develop multiple use cases at once, whether or not to outsource development, what budget is available, and, most importantly, what organizational capacity is available for the delivery squad, the central team that will deliver the use case. Whom to

include in the delivery squad and how to manage it are vitally important to determine early on.

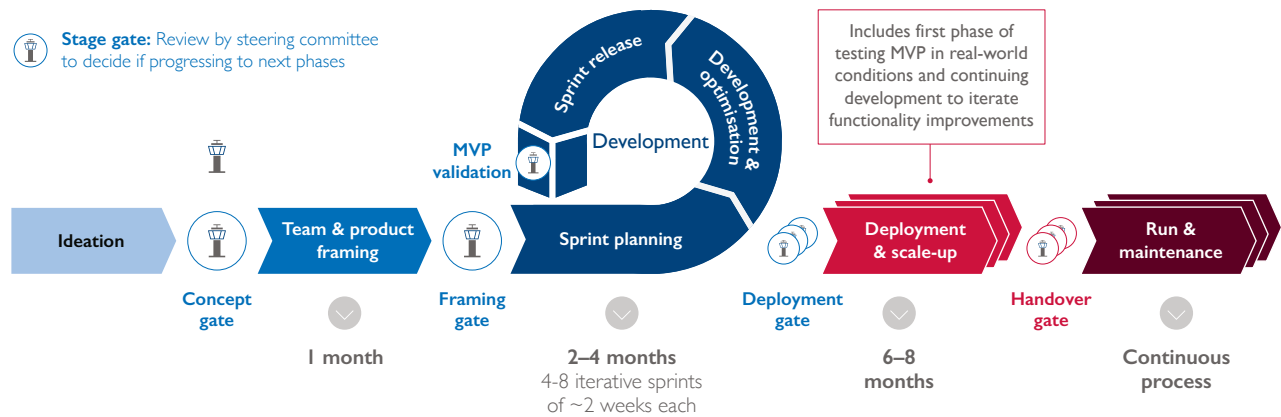
The delivery squad will include digital roles, such as data scientists, data engineers, software engineers, and developers, who will build the product. Three key roles from the organization also need to be represented:

- **Product owner:** Defines the purpose of the product and manages the squad, and needs to be fully dedicated for the team and product framing and development phases.
- **User representative(s):** Provides input on the user experience and supports the product owner in ensuring delivery, and needs to be fully dedicated for the team and product framing and development phases.
- **Subject matter expert(s):** Topic expert dedicating part of his or her time to assist the squad in developing the product.

Another important consideration is the role of the stage gates that separate the phases (see Figure 20). These are decision points for the project steering committee to decide whether or not the project should progress to the next phase. There is no initial guarantee that a use case idea will be successful. At every phase, progress should be reevaluated against priorities. This is also why having a portfolio of high-priority use cases is beneficial. Working with a “fail-fast” mentality, a decision can be made early on to pivot to a more promising use case and limit losses from continuing with an idea that is not going to work.

A select few of the most important questions to address for each phase are detailed in the following bulleted series. These questions should be addressed upfront to the fullest extent possible to facilitate planning. Note that our study (included in previous sections of this report) covered the ideation phase, with some initial inroads made to product framing.

Figure 20: Use case delivery model phases



Ideation:

- What are the pain points in the HIV landscape that we are trying to solve with our use cases?
- Which stakeholders could provide the best input to understand those pain points?
- What is the scope of our use case ideation?
- What criteria should we use to prioritize a long list of use case ideas?
- How much impact will the deployed and scaled use case create (high level)?
- How feasible will the use case be to deploy and scale (high level)?
- What are the “deal breakers” to decide whether or not to proceed with a use case at each stage gate?
- Which stakeholders should be involved in the stage gate decisions and how?

Team and product framing:

- How do we enable partners to drive the use cases forward with us?
- How should we choose a digital partner for implementation (if required)?
- Which stakeholders will be represented in the delivery squad, and at what level (for example, government, private sector, NGOs)?
- What will the governance model be for the delivery of the project?

- How do we retain management of the use case development while being cooperative with partners?
- What does the roadmap to developing a minimum viable product (MVP) look like (high-level)?

Development:

- How should the development of requirements for the MVP be split into sprints?
- What is the “soft launch” plan for the MVP (where will it be tested and by whom)?

Deployment and scale-up:

- What is the deployment plan? What does full deployment and scale look like?
- At what point do we progress from testing the MVP in the soft launch to full deployment and scale-up?
- What are the key challenges expected for deployment at scale?
- How do we involve diverse stakeholders in addressing these challenges?

Running and maintenance (ongoing process):

- How do we ensure that the use case is sustainable in the long run?
- Which stakeholder(s) will be involved in the running of the use case?
- Which stakeholders will pay for the ongoing running costs (minimal)?

- Which stakeholder(s) will “own” the use case once launched?

Due to intensive capacity and other requirements, most organizations will be able to develop at most two to three use cases within a financial year. The selection of these two to three use cases should be informed by the

impact and feasibility assessment of the shortlisted use cases, but also by your organization’s specific priorities, objectives, and constraints. This also depends on available budget, with an understanding that not all use cases in the framing phase are likely to reach the scaled-up deployment phase.

8. Conclusion

The purpose of this report is to provide a launchpad for those looking to use advanced analytics in the global fight against HIV/AIDS. As we work toward the goal of 95-95-95 by 2030, finding more effective, rapid, and cost-efficient ways of testing PLHIV, getting them onto treatment, and keeping them stable is more important than ever. This is particularly relevant at the time of writing, as COVID-19 imposes additional constraints on PLHIV receiving quality care, while underscoring the potential opportunity to leverage advanced analytics to make our service more dynamic and responsive, such as improved patient tracking and focusing/tailoring interventions efficiently to hotspots of need.

Data is foundational to advanced analytics and ensuring its quality and richness should be two key priorities for those working in this space within the HIV landscape, whether through improving patient-level data collection and management practices as described in Section 6.1 or by incorporating external data sources with insights on the populations you serve. Improving quality data availability will help to ensure the success of the ten top use cases detailed in this report, which leverage data to provide insight into the problems faced in HIV service delivery today.

Potential advanced analytics use cases suggested in this report range from identifying and locating people at risk of being infected with HIV to transferring stable patients on treatment into less resource-intensive treatment models. The highest-priority use cases for your organization or country will depend on existing priorities, programs, and progress toward 95-95-95, but whichever use cases

apply best to your context, their impact in accelerating improvements to clinical outcomes and cost efficiency are likely to be significant. For example, in contexts where prevention and diagnosis are significant challenges in the journey toward achieving the first 95, use cases such as the High-Risk Hotspotter are estimated to provide reductions in new infections of 3% to 7% per year and increases in diagnosis rates of 6% to 11%. Leading with use cases such as the Better Booker in countries where long waiting times have a profoundly negative impact on treatment adherence could result in 8% to 12% less patients falling off treatment each year. Optimizing the use of decentralized care models to reduce the number of facility visits through use cases such as the Decongestion Driver could result in savings of 50% to 80% in expected costs per patient.

We expect the learnings taken from this report will be further developed by those working in other contexts and by those who take the use case designs into further stages of delivery, such as development and deployment. Taking an advanced analytics use case from an exciting opportunity to an impact-generating reality is challenging at the best of times. However, the true value of implementing effective digital solutions can be seen at the “worst of times.” The current global COVID-19 pandemic is a poignant example of this, where data-driven solutions such as using geospatial analysis for contact tracing have proven to have profound effects on outbreak prediction and containment. Looking ahead, it is vital that stakeholders across the value chain join together in communities of practice to develop and implement use cases to improve HIV service delivery and hasten progress toward 95-95-95.

9. Appendix

A. Glossary

AGYW	Adolescent girls and young women
ART	Antiretroviral therapy
ARV	Antiretroviral
BCG	Boston Consulting Group
CCMDD	Chronic medication dispensing and distribution
CHW	Community health worker
CM	Case manager
COVID-19	Coronavirus disease of 2019
FM	Facility manager
HCW	Health care worker
HR	Human resources
ID	Identification
IoT	Internet of Things
IT	Information technology
KPI	Key performance indicator
LMIC	Low- and middle-income countries
LTFU	Loss to follow-up
MER	Monitoring, evaluation, and reporting
MMC	Medical male circumcision
MVP	Minimum viable product
NGO	Nongovernmental organization
PEPFAR	The U.S. President's Emergency Plan for AIDS Relief
PHC	Primary Health Care
PLHIV	People living with HIV
PM	Program manager
PrEP	Pre-exposure prophylaxis
ROI	Return on investment
SMS	Short message service
STI	Sexually transmitted infection
TB	Tuberculosis
UNAIDS	Joint United Nations Programme on HIV/AIDS
USAID	United States Agency for International Development

B. Long list of pain points

Note that the below pain points are the expressed concerns and perspectives of the various stakeholders interviewed in this research, across funders, implementing NGOs, public health clinicians, supply chain managers, etc.

Prevention (e.g., MMC/PrEP):

- Difficult to tailor prevention programs at vulnerable populations (e.g. AGYW) as they are larger and more general
- Difficult to determine high-risk individuals within a geography and understand their observable characteristics without asking direct questions
- Side effects of PrEP, including nausea, abdominal pain, headache
- People don't know about PrEP
- People are bad at understanding their own risk and can't identify "high-risk" periods of sexual activity
- Unable to track who has been treated for an STD in a public facility, and this is a strong indicator of who should be taking PrEP
- If you don't take PrEP properly, with a ramp-down period, you could build ARV resistance
- Permission from husband/partner required by a woman who wants to take PrEP
- Stigma around being on PrEP, notably from intimate partner, such as it implies high-risk sexual behavior (the "promiscuity pill")
- Stigma around collecting PrEP medication, such as "HIV medication" from a clinic, for example, in distinguishable packaging brown bag (same as ARVs)
- PrEP-specific stigma from HCW who think patient are promiscuous
- Taking a daily pill is not discrete; very difficult to hide it
- PrEP requires regular clinic visits to collect prescription and perform HIV test (every three months), as well as liver function checks every six months
- No integrated services available (i.e., PrEP as part of holistic health care services)

- Treatment fatigue: effective use requires daily adherence, which is inconvenient
- Ambivalence about taking a pill for prevention versus treatment: attitude that a person can take a pill now (PrEP) or later (ARV)
- No differentiated treatment models available for taking PrEP (e.g., external pick-up points), regardless of the health status of the person taking it
- Difficult to manage quality control in private sector-led MMC programs, including fraud and quality of care

Case finding:

- Struggle to message correctly to the market to target specific populations
- Not able to quickly respond to community sentiment on a particular issue
- Don't know in which areas PLHIV who have not been tested are concentrated
- Difficult to segment and find the PLHIV who fall outside of key populations
- Cannot measure or track the reduction of stigma from community outreach/marketing
- Don't understand reasons why HIV is still being transmitted vertically to infants
- Difficult to keep CHWs up to date with constant training
- Giving CHWs smartphones makes them a target for crime in communities
- Potential threats to the safety of CHWs (e.g. burglary, access issues) hinders ability to go to certain hotspots
- Adding more eTools for CHWs could detract from existing work, as they already have a long list of variables they have to check for
- Unable to get young people to test/to get on treatment; need to incentivize more (e.g., free Wi-Fi at clinics)

Testing:

- Risk profiling of patients entering facilities for non-HIV services is not done efficiently

Figure 2I: Some major pain points along the patient journey

NONEXHAUSTIVE | SIMPLIFIED

Overview of major pain points in HIV service delivery

Prevention	1st 95 Diagnosis			2nd 95 Treatment			3rd 95 Viral Suppression	
e.g., MMC/PrEP	Case finding	Testing	Counseling	Linkage	Early treatment	Stable treatment	Viral suppression	Differentiated models
Difficult to tailor programs at vulnerable populations as they are larger and more general	Don't know in which areas PLHIV who have not been tested are concentrated	Risk profiling for patients entering facilities for non-HIV services is not done efficiently	Counselors can't identify PLHIV that will struggle to stay on treatment or which factors will influence this	"One-size-fits-all" approach to test-to-treat, not taking people's specific requirements into account	5 repeat visits required in first 2-3 months of treatment - frequency leads to stigma/missing too much work	Ineffective patient tracking in case management (e.g., across different health databases, facilities, cities, etc.)	An insufficient share of stable PLHIV are tested for viral suppression	Lack of clarity on the reasons for medicine parcels not collected from external pick-up points (PuPs)
Difficult to determine high-risk individuals within a geography and understand their observable characteristics	Don't know what to expect regarding HIV testing outcomes for particular segments/ locations	Don't know what to expect regarding HIV testing outcomes for particular segments / locations	Can't differentiate between patients ready for same-day initiation and those needing more counseling	Following up with patients after testing positive to ensure that they have been linked to treatment is not done rigorously	Difficult to navigate multiple touchpoints (testing, treatment, different providers), waiting times, and inconvenient hours	Important patient or civil society-led information not fed back into facility decision-making	Many patients that are failing on second-line treatment can't be moved to third-line due to cumbersome process	Dropout from differentiated models caused by stable patients needing to return to facilities several times a year
No integrated services available (e.g., PrEP as part of holistic Health care services)	Struggle to message correctly to the market to target specific populations	Delay between testing and results can lead to losing track of patient	Men often feel uncomfortable being counseled by women nurses	Fear of status disclosure in facilities (e.g., HIV-specific files and stationary, dedicated waiting lines, etc.)	Professional, friendly service and clinic space, as well as treatment support, often not meeting patient expectations	Patients feel symptoms improve or can't afford "holistic" approach (e.g., healthy food) so give up on treatment		Identification of specific populations to target for differentiated treatment models is difficult
		Index testing yield is too low and can sometimes lead to intimate partner violence if not handled well	Some PLHIV need more counseling than others and this is not taken into account when allocating staff time	Data management: <ul style="list-style-type: none"> Data collection is manual, time-consuming, and, in some systems, dependent on data Internet connectivity Subpar filing systems lead to duplication of information and difficulty in obtaining patient files Lack of agreement and motivation between stakeholders to integrate various health systems and databases Operations management: <ul style="list-style-type: none"> Difficult to manage "bullwhip effect" in supply chain caused by policy changes, clinician training (e.g., on new regimen), private-public crossover, migrant stockpiling, seasonal cross-country migration, etc. Difficult to know when to intervene in a program to ensure targets are met 				
Staff: <ul style="list-style-type: none"> Not clear what is the right balance of staff roles for best outcomes at most efficient budget—staffing levels not responsive to changing demand Legacy staffing job specifications are not fit for purpose, and it is not clear that roles fit needs (e.g., pharmacist running warehouse) Importance of roles that directly interact with patients (e.g., community health workers, counselors, etc.) is not emphasized enough 								

- Index testing yield is too low
- Index testing can lead to intimate partner violence (IPV), for example, if a woman is the index and her husband is tested because of her
- Don't know what to expect regarding HIV testing outcomes for particular segments/locations
- Test-to-treat leading to some patients not being ready for treatment (require more counseling and more than one touch point)
- Delay between testing and results can lead to losing track of patient
- Don't know what testing information has been shared with and understood by patients

Counseling:

- Counselors can't identify PLHIV who will struggle to stay on treatment, or which factors influence this response
- Some PLHIV need more counseling than others and this is not taken into account when allocating staff time
- Can't differentiate between patients that are ready for same-day initiation and those that need more counseling/other interventions
- Men often feel uncomfortable being counseled by women nurses

Linkage:

- Lower share of men on ART versus women and children, lost predominantly at initiation step
- Difficult to target certain populations for care (e.g., young men as cultural barriers and gender roles lead men to not seek treatment)
- Don't know which PLHIV who have tested positive are unlikely to show up at a clinic for linkage, or what timing/location would encourage this
- Required steps to take between testing positive and getting on to treatment aren't well understood
- PLHIV think they have other options (e.g., traditional healers, prophets)
- "One-size-fits-all" approach to test-to-treat does not take people's specific requirements into account (e.g., days they can come to clinic)

- Following up with patients after testing positive to ensure that they have been linked to treatment is not done rigorously
- PLHIV don't feel safe to admit that they have previously been on treatment for multiple reasons (e.g., fear of judgment from nurses, etc.)
- Losing track of patients after first contact, due to reasons such as when their phone numbers change
- Lack of data sharing between facility and linkage officer to check if a missed linkage appointment is the result of a patient moving
- Financially dependent women are fearful of their partners finding out their status so they don't seek treatment
- Fear of status disclosure in facilities (e.g., HIV-specific files and stationary, color-coded HIV files, dedicated waiting lines, etc.)

Early treatment:

- Multiple repeat visits required in first two to three months of treatment—frequency leads to stigma (perception that they "must have HIV") or being fired for missing so much work
- Difficult to navigate multiple touchpoints (testing, treat, different providers, etc.) in first few months of treatment
- Lack of overall case management during intensive clinic period
- Of patients not on treatment, majority are lost within first six months
- Delay in flagging LTFU if treatment data is not captured immediately in the system
- Long lines and waiting times
- Don't know what proportion of patients are coming on time to their appointments
- Inconvenient opening hours (no late/weekend hours), leading to low retention
- Unable to track patient mobility and its impact on treatment outcomes
- Adjusting treatment regimens to manage side effects is not done efficiently

- Insufficient treatment support (e.g., around side effects—notably TB)
- Lack of patient support (e.g., support groups)
- Clinics unofficially close earlier than they are supposed to, and workers often leave early
- Men are not comfortable being vulnerable with female nurses and counselors and there are insufficient male clinicians for male patients
- Stigma continues (e.g., PMTCT: mother can justify frequent clinic visits while pregnant and can get ARVs, but difficult to do so once breastfeeding because of stigma)
- Unfriendly/unprofessional in-facility environment and HCWs
- Stigmatized waiting areas
- Young people still face stigma and judgment from nurses
- Poor in-facility service (e.g., lost test results)
- Poor initial health (e.g., low CD4 count)
- Don't know how to implement retention programs at scale
- There isn't a retention officer in every facility
- Social media campaigns need to emphasize that treatment interruption can happen, but it's important to go back
- Men do not feel comfortable in clinics "designed for women," (e.g., lots of crying children, female staff, etc.)
- Poor data on patient opinions and preferences and how this impacts retention
- No data visibility on what's happening in treatment in smaller catchment areas
- Social/lifestyle issues lead to drop-off (e.g., alcohol abuse, poverty, violence) and can't foresee when they will be a challenge for a particular patient
- Vertical HIV programs and data do not include the holistic care of patient required for true quality and impact
- Patient can't afford "holistic" approach (e.g., healthy food) so give up on treatment
- Currently not capturing or focusing enough on patient-related outcomes that matter for quality of care (e.g., number of people who have died)
- Important patient or civil society led information not fed back into facility decision-making
- Stockouts are an issue and on-shelf availability does not incorporate actual treatment/dispensing data and other predictive information (e.g., policy changes, treatment guideline update training schedules, patient migration, community behavior)
- Challenging to get patients to stay on treatment long term/forever because of general treatment fatigue
- PLHIV with a high CD4 count (i.e., that are well) not on treatment because of fear and stigma around being on ART

Stable treatment:

- Ineffective patient tracking in case management (e.g., can't track across different health system databases or areas)
- A large share of recorded LTFU is not "real" LTFU: patients have moved and are receiving treatment at a different facility
- Patients are resisting the switch to TLD
- Providers are failing to transition patients to TLD

Viral suppression:

- An insufficient share of stable PLHIV are tested for viral suppression
- Many patients who are failing on second-line treatment can't be moved to third line due to cumbersome process

Decentralized medicine distribution:

- Don't know the reasons for medicine parcels not collected from external pick-up points (PuPs),
- Lag in flagging of missed pick-ups causing operational expense/difficulty
- Suboptimal decision-making of where to locate external PuPs to be most effective

- PuPs are sometimes not economically sustainable in the long run because of varying range of patients collecting medicine at each one
- Insufficient number of PuPs
- Dropout from differentiated treatment models caused by long-term stable patients needing to return to facilities several times a year
- Insufficient number of patients served through differentiated treatment models to reach 95-95-95 targets by decongesting clinics
- Clinicians not transferring patients to differentiated treatment even if eligible
- Identification of specific populations to target for differentiated treatment programs is difficult
- Typical couriering issues experienced (e.g., incorrect address, patient details, medication)
- Large portion of recorded LTFU in facility is not “real” LTFU: some patients are picking up medication at external PuPs but this is not reflected in treatment databases
- Getting patients sufficiently stable to enter differentiated treatment programs is a challenge
- Concerns around on-selling of drugs and drug abuse if PLHIV are transferred to differentiated treatment programs
- Not possible to tell if patients are living or working near facility (i.e., cannot track pick-up behavior based on employment or daily travel)

Cross-cutting – Data management:

- Treatment data is not timeously recorded because there aren’t enough data capturers at facility level
- High number of insufficiently trained data capturers
- Collection of data is manual, time-consuming, and, in some systems, dependent on Internet connectivity—a struggle in more rural areas
- Poor management and oversight of data capturers
- Lack of enthusiasm around analytics because of “data overload”
- NGOs invest in complicated analytics when focus

should be on data integration and quality, and simple analysis that can yield effective results

- Unable to tie staff data to indicators (i.e. patient outcomes)
- ID fraud/duplications means that ID number is not always a unique identifier
- Currently don’t have data on consumption to make a call on whether stock levels are sufficient at a facility
- Poor job currently being done of verifying cellphone numbers in facilities which presents challenges for reaching patients
- Shortage of data scientists and other data skills in public health system
- Need to integrate patient-facing mobile applications with patient records and treatment data to enable authentication & provide more personalized services
- Lack of agreement and motivation between stakeholders to integrate various health systems and databases
- Low trust in biometrics so patients don’t want their fingerprint used as an identifier
- Most PHC facilities don’t have access to online databases
- Stakeholders are tracking similar outcomes, but using different metrics
- Sub-par filing systems lead to duplication of information and difficulty in obtaining patient files
- Staff are often not “tech savvy,” so can’t draw basic insights from data
- Legal concern regarding data access permissions for analytics
- Mishandling of patient files (e.g., files of high viral load patients are put aside and sometimes get misplaced or are not reported)

Cross-cutting – Operations management:

- Challenges in working with community-based organizations that don’t have good practices and processes, and are susceptible to fraud and mismanagement of funds

- Difficult to know when to intervene in a program to ensure targets are met
- Difficult to manage “bullwhip effect” in supply chain caused by policy changes, clinician training (e.g., on new regimen), private public crossover, migrant stockpiling, seasonal cross-country migration, etc.
- Unable to redirect supply between facilities if there’s shortage/excess stock, and unable to redirect delivery of supply flexibly
- Abnormalities/fraudulent behavior in stock procurement processes currently not being identified
- Changing regimen prescriptions by clinicians poses threat to supply chain—need to be adequately tracked and significant changes predicted ahead of time, including general connection of patient clinical demand to supply forecasting
- Significant money spent on nontendered purchases resulting from inability to manage suppliers effectively
- Supplier performance is not adequately tracked and monitored, and penalties for noncompliance with tender not enforced transparently (or at all)
- Medicine supply-side projections are not robust and suppliers not held accountable
- Need to disaggregate purchases to facility level to improve visibility and reduce bad practices currently fostered by centralized environment
- Coordination of program monitoring at management levels needs to filter down to facility level
- mHealth can be less effective in rural areas due to poor Internet connectivity and low use of smartphones/tablets
- Transition of some patients to TLD and others remaining on TLE (and uncertainty around clinical effectiveness and supply availability of the new regimen) will further complicate supply and demand forecasting
- Legacy staffing job specifications are not fit-for-purpose and it is not clear that roles fit needs
- Poor measurement/tracking of staff performance
- Don’t know what HCWs are actually doing with their time
- Clinics are underresourced (e.g., clinics have one phone so only the manager can call people to follow up)
- Some nurses do not always have a strong work ethic and do not listen to facility managers
- Some workers who directly interact with patients (e.g., community health workers, counselors, etc.) don’t realize the importance of their jobs
- Facility managers are not capacitated to do their jobs, causing dysfunction in the health system
- Facility managers can’t always implement performance management due to attitudes, labor unions, etc.
- There’s a shortage of skilled staff—need to bring more academic resources into the sub-national health systems
- No leadership in public facilities (i.e., not all facilities have a manager/poor leadership by facility managers)
- Environments of distrust between facility managers and staff (e.g., clinic managers don’t want staff to be able to use the telephone because of fear of mismanagement)

C. Long list of use cases

Prevention (e.g., MMC/PrEP):

- As a program manager (PM), I want to identify more granular sub-segments (and their characteristics) and geographic hotspots of high-risk populations to enable tailoring of targeted prevention program
- As a PM, I want to predict the likelihood of a patient joining or dropping off of PrEP to tailor communication/level of effort to the patient
- As a clinician, I want to predict the likelihood (and type, if possible) of an individual experiencing side effects on PrEP to enable tailoring of regimen

Cross-cutting – Staff:

- Workforce reduction potentially looming (as a result of budget cuts), but don’t know optimal mix of various roles or where to cut

Case finding:

- As a PM/CHW/case manager (CM), I want to optimize community messaging to tailor better response, based on sentiment analysis and/or population behavior data (e.g., retail shopping data, complaints at facility level)
- As a PM, I want to identify geographic hot spots of undiagnosed/untreated PLHIV to optimize deployment of CHW, launch targeted awareness campaigns, outreach, etc., and adjust near real time with population changes, including safety issues
- As a PM, I want to predict the likely impact of awareness campaigns and measure actual attributable impact better (e.g., stigma sentiment analysis) in order to decide which campaigns to fund

Testing:

- As a health care worker (HCW), I want to predict which patients presenting in facility are high risk for HIV to ensure they are targeted for testing
- As a CM, I want to predict effectiveness and risk of index testing for an individual and optimize outreach to increase index rate/yield

Counseling:

- As a PM/facility manager (FM), I want to predict which recently diagnosed patients will be at risk of LTFU in order to optimize allocation of best counselors/case managers to them, including matching men to men optimally
- As a PM/FM, I want to monitor performance of counselors to target trainings at poor performers and optimize time spent per patient
- As a FM, I want to optimize time spent with the client for CHWs, counselors, and nurses/clinicians for greatest impact and workforce efficiency

Linkage:

- As a CM/CHW, I want to predict the likelihood of a newly diagnosed patient not initiating treatment and identify the key areas of support they need, to tailor interventions (e.g., personalize nudges, optimize appointment schedule/location, test-to-treat readiness)
- As a CM, I want to tailor the test-to-treat journey to each patient based on channel and scheduling likely to be most successful

Early treatment:

- As a CHW/CM/FM/PM, I want to predict likelihood of a patient falling off treatment and the key areas they need support on to be able to target interventions (e.g., personalized case management and communications)
- As a PM, I want to be able to predict patient-level treatment data indicators and compare to other relevant data sources to make up for delays in data capturing and other data quality challenges that skew results and affect grant allocation
- As a clinician/FM, I want to minimize facility waiting time and improve overall patient experience with targeted support based on the facility capacity, type of appointment (e.g., initial HIV test, medication pick-up, clinician appointment), patient risk, and behavior
- As a CM, I want to predict the progression of HIV based on a patient's profile and point of diagnosis, in order to help the patient understand their condition, and better tailor "care interventions" over the patient lifecycle
- As a PM/FM, I want to have an up-to-date and predictive dashboard with info on facility and nearby population issues so I can get ahead of challenges and provide a quality first (and lasting) experience for patients (e.g., population behavior trends, community shocks, population migration and real-time traffic info vs. related staff capacity, early/predictive stockout warnings, responsiveness to complaints)
- As a clinician, I want to optimize prescribed treatment regimens based on, for example, likelihood of side effects of different treatment regimens, other risk factors (i.e., co-morbidity), patient behavior

Stable treatment:

- As a CHW/CM, I want to have an integrated view of my patients across all health data systems to have a holistic understanding of their treatment and provide the necessary personalized care
- As a CM/PM, I want to optimize tracking of patient mobility (e.g., between facilities, cities, provinces) to not miscategorize LTFU, predict capacity needs, provide uninterrupted support, enable efficient referral processes, etc.
- As a clinician/FM, I want to predict which stable patient profile(s) will be resistant to changing

regimen (e.g., not transition to new TLD regimen) to target nudges and alert clinicians to patients' potential reluctance

Viral suppression:

- As a PM/FM/clinician, I want to identify/predict patients that are not taking sufficient viral load tests and develop an action plan to improve testing rates based on their needs
- As a clinician/FM, I want to predict the future outcome of a patient's viral load test and preemptively start to target patients for differentiated treatment programs or test patients with consecutively high viral loads for drug resistance

Decentralized medicine distribution:

- As a PM/CM, I want to predict the likelihood of a patient not picking up their medicine at an external PuP to preemptively target interventions and reduce returns in supply chain
- As a PM, I want to optimize the location/number of external PuPs to serve the greatest number of patients on decentralized models (current and future) while allowing greatest reasonable volume per PuP for their business model
- As a CM, I want to track when a patient on a differentiated treatment program will have a public-sector facility touchpoint and tailor timing of interventions to minimize LTFU
- As a PM, I want to identify geographical/facility hotspots of potential candidates for differentiated treatment programs in order to invest resources to shift more people in those areas to decentralized models, including tracking of daily travel (e.g., due to employment)

Cross-cutting – Data management:

- As a facility manager/program manager, I want to automate data ingestion using natural language processing to optimize workforce needs, reduce spend, and improve data quality
- As a PM/FM, I want to monitor performance of data capturers to tailor targeted trainings at poor performers and optimize allocation/timing of data capturing and cleaning

Cross-cutting – Operations management:

- As a funder/PM, I want to be able to quickly detect abnormal/fraudulent transactions by community-based partner organizations

- As a funder, I want to track program outcomes versus targets in relation to population trends and facility issues, and receive directive alerts when the program is falling behind so that I can target interventions
- As a FM/PM, I want to improve/automate data ingestion of invoices in order to accurately track costs and optimize spend at a granular level
- As a FM/supply chain manager (SCM), I want to automate ingestion of stock data to help, for example, integrate medicine dispensing data at a patient level with inventory data
- As a FM/SCM, I want to predict stock requirements at a facility level based on actual and forecasted patient treatment data and population trends
- As a PM/SCM, I want to detect abnormal/fraudulent supplier behavior in order to enforce compliance measures, save money from nontender purchases, and direct volumes to suppliers that are performing well
- As a PM, I want to optimize the decision between direct delivery and depot model (e.g., per facility, supplier, delivery), considering the types and volumes of SKUs required per facility, location of the manufacturer, patient demand variability, etc.

Cross-cutting – Staff:

- As a FM/PM, I want to optimize the amount of time spent by staff on each activity to improve scheduling and maximize efficiency
- As a FM/PM, I want to optimize staff allocation across facilities and days/times to match need and optimize workforce allocation for each facility based on impact and changing capacity needs, including predicted demand variability

D. Detailed descriptions of impact and feasibility sub-criteria

Impact: Sub-criteria

Clinical outcomes impact was assessed by the impact of the fully deployed use case on HIV prevention and treatment at the patient level, including the likely number of PLHIV impacted along the patient journey and/or the improvement to quality of care received. To measure the impact, we used key performance indicators (KPIs) from stakeholders' monitoring and evaluation strategies and stated government priorities. These impact metrics included the number of positive HIV tests and the

number of patients on treatment. For each use case, the definition of PLHIV impacted differed based on its focus areas and objectives, but these were made comparable wherever possible.

Cost-efficiency impact was assessed by the ability of the use case to reduce spend in the HIV landscape, measured by the size of the cost bucket affected and the extent and likelihood of cost savings from the fully deployed use case.

Clinical outcomes comprised 70% of the total impact score, with cost efficiency comprising the remaining 30%. This is because the ultimate objective of cost savings would be to treat current PLHIV more cost-effectively, enabling additional PLHIV to be treated—circling back to clinical outcomes. Cost improvement with no benefits to clinical outcomes was not the objective of this piece of work.

Feasibility: Sub-criteria

Change management and sustainability was assessed by considering the number of stakeholders involved in operating the fully deployed use case, and how difficult it would be for these stakeholders to change entrenched ways of working. An additional consideration was how sustainable the use case would be to run without continuous donor funding and management.

Data condition was assessed by looking at the access to, availability (number of data points available, frequency of availability) of, and quality (cleanliness, robustness) of the “must have” data sets for the use case. A “must have” data set is one that, without which, the use case would not achieve even minimal level of its potential impact. Other data sets can be “nice to have”—they would add richness, but are not a fundamental requirement.

Data science feasibility was assessed by looking for examples of similar use cases to determine if accurately predicting or optimizing the target variable(s) had been done before. We also looked at the complexity of the advanced analytic technique(s) to be used. However, it was difficult to assess this criterion without a proof of concept or at least access to a data sample. The data science behind a use case may be feasible theoretically, but without sufficient data of sufficient quality, it can be impossible.

IT/Technology requirements include the hardware and software required to build and run the use case, and the complexity thereof. This includes Internet connectivity, data lakes or warehouses, and fit-for-purpose apps or other tools required.

It is widely accepted that change management is one of the most important levers—and most difficult—to get right when implementing a use case, and for this reason it comprised 40% of the total feasibility score. For many LMICs, data condition is a close second due to access, availability, and quality issues explored in this report. For this reason, data condition comprised 30% of the total feasibility score. Data science feasibility and IT/technology requirements made up 15% each to complete the weighting—most of our use cases follow tried and tested advanced analytics approaches and IT/technology requirements across the use cases were quite similar, with no particularly onerous requirements.

E. Categories and details of external data sources

Non-health population behavior data

a) Socio-demographic data

Socio-demographic data sources provide powerful insights into the social and demographic makeup of a population and some of the structural and societal influences behind the way people in an area or population group behave. This category has a wide variety of useful variables including age, gender, employment status, income, purchasing power, marital status, household size, access to water and sanitation, alcohol abuse statistics, media consumption, and literacy.

Socio-demographic data has had widespread application in public health globally, including current use as a source to predict transmission risk of infectious diseases such as COVID-19¹ at a granular area level, up to 1 km² resolution in some instances. Some high-quality socio-demographic data providers make this information available on geospatial maps, enabling precise derivation of insights for an area, and making it possible to overlay multiple data layers to gain richer understanding of an area's profile.

For our top use cases, socio-demographic data can provide a range of critical insights, such as understanding the profiles of at-risk populations for the High-Risk Hotspotter and the Risk Rater, providing rich predictive risk factors of falling off treatment for the Fallout Forecaster, and understanding social and demographic factors that cause population mobility for the Mobility Monitor and the Stock Saver.

b) Consumer retail data

This category of data provides insights into consumer shopping behavior and trends of different population groups and/or consumer behavior in a specific geographic location. Variables that could be beneficial

in this category include location of stores, frequency of visits by customers, average basket size of specific customers or in specific areas, time of day and day of week of highest traffic volume, location of primary store from a customer's home and work, etc. The profile of the population served by different retailers is often linked to income and affluence, and while some of the identified providers are perceived to serve higher income brackets, some overlap in target populations is likely in these retailers as well.

Engagement with multiple providers of consumer retail data indicated that data from loyalty or rewards programs provides the deepest insights into population behavior. Loyalty card data is commonly used to tailor promotions and channel recommendations for consumers, but the ability to monitor the value, frequency, and location of purchases, and what rewards customers respond to, gives this type of data considerable application potential. Depending on maturity and success of these loyalty programs, data sets identified can contain millions of customers.

Consumer data could be leveraged in our use cases such as the Fallout Forecaster and the Risk Rater to identify risky behavior and potential community financial trends that may be predictive of challenges to stay on treatment. Consumer data could also elevate the predictive ability of the Mobility Monitor by identifying common corridors of movement by customers across income categories at different times of the year.

c) Social media and online news data

Social listening and analysis of posts on public social media platforms and online news and forums provide great insights into the general volume and sentiment of chatter around specific topics. With the growth in number of platforms and people making use of those platforms globally, the data generated is growing exponentially and analysis of this data for multiple applications is on the rise. For example, social media data has been used to monitor the spread of COVID-19 based on tracking the self-reporting of symptoms². Millions of people in LMIC—mainly concentrated in big cities—use social media, such as Facebook, Twitter, and Instagram. Stigmatized topics tend to generate lower volumes of chatter, and consideration needs to be given to the key words used in sentiment analysis due to the variety of languages spoken in the country, to ensure as much relevant chatter as possible is captured.

Sentiment analysis could be used to identify issues raised about particular facilities that can be useful for use cases such as the Staff Supporter and the Better Booker. It could also provide insight into area-specific events that

could have an adverse impact on community health workers to enable alerts for the High-Risk Hotspotter and the LTFU Locator.

d) Telecommunications data

The true power of the data collected by mobile network operators is their access to geolocation information on individual mobile devices, enabling them to track mobility of the population. Countries with high smartphone penetration across income groups coupled with regulatory requirements for each SIM card to be registered to an individual make this a rich source of data, with more stringent and sensitive access requirements. Increasingly, we are seeing mobile network operators play an important role in using geolocation data for contact tracing in epidemiological outbreak response across the globe.

Telecommunications data could provide game-changing insight for the Mobility Monitor, enabling more precise and timely tracking and prediction of patient mobility. Geolocation data could also be leveraged to understand facility demand and average time spent by individuals in a facility to inform optimization of patient and staff time in the Better Booker and the Staff Supporter.

e) Financial services data

Financial services data sources provide insight into consumer trends and attitudes toward financial services across banking, insurance, savings, and credit products. A consumer's income level and socio-demographic factors play a critical role in their selection of financial services products, and the manner in which they manage these products provides insights into the profile of a consumer or area. For example, lower credit scores have been identified as markers of high disease burden in some contexts³, and increases in insurance claims, including simple products like funeral insurance, can help predict when financial hardship is hitting communities.

For our use cases, financial data could add richness to the identification and prediction of at-risk patients in the Risk Rater and is particularly insightful for the Fall-Out Forecaster, as financial hardship is a common shock that complicates treatment for patients by, for example, rendering them unable to afford transport to a facility.

Health-related population behavior data

a) Private-sector medical records

Private medical insurers collect rich data on their membership base, including variables such as patient demographic information, treatment and claims history, facility usage, treatment regimen, etc. While the vast majority of PLHIV in LMICs receive treatment from the

public sector, private-sector health records could still provide a wealth of information on those HIV patients who are in the private sector, and insights to enable comparisons for greater understanding of the factors that drive differences in outcomes, for example, treatment adherence rates based on demographic factors. The system also experiences mobility of patients between the public and private sector largely driven by financial constraints forcing patients to seek cheaper options when they exhaust medical aid benefits or are no longer able to afford the service. This movement and its impact on public-sector outcomes, including unexplained loss of patients from treatment, is currently not tracked and could be better understood through this data.

A number of private medical insurers already use patient data for advanced analytics applications internally, including predicting risk of chronic illness and proactively targeting interventions at at-risk patients. For our use cases, this data could provide information on comorbidities of PLHIV as input into the Risk Rater and drivers of adherence on disease management programs to add richness to the Fallout Forecaster. The data could also enable comparison of drops and upticks in private-sector treatment numbers compared with trends in the public sector to add richness to the Treatment Data Triangulator.

b) Private pharmacy data

Data syndication platforms and providers of data science capabilities have been identified as key sources of good-quality aggregated data on private pharmacies and medicine dispensing. The data available to these players depends on their offering and can include variables such as the volume of over-the-counter and prescription products sold in specific pharmacies, customer scripting data, profiles of customers on chronic medication, etc. De-identified transaction data is often leveraged by pharmaceutical manufacturers for business intelligence into performance of their drugs compared with other brands. Customer-level data has been used in some contexts for monitoring and promoting treatment adherence by pharmacies to ensure customers stay on treatment and continue to use their outlets. Some providers of data analytics on pharmacy customers have the ability to track customers across stores and brands in their network, providing powerful insights into patient mobility and behavior of individuals on chronic treatment.

For the top ten use cases, this category of data can be leveraged for adding richness to patient risk prediction in the Fallout Forecaster and the Risk Rater, and for providing insights into bullwhip effects from public-private crossover that could affect the demand for public-sector antiretroviral medication in the Stock Saver.

c) Private-sector facility and laboratory data

Private hospital groups and treatment centers can provide some information regarding in-facility patients. However, the data is usually mostly focused on hospital admittance and patient referrals rather than the chronic primary-care treatment used for HIV, and there is limited overlap in the facility patients and the HIV-positive population. Pathology data from private-sector laboratories can provide useful insights into aspects, including the burden of disease in certain areas, trajectory of incidence, and the profile of individuals most likely to test positive for certain conditions. While the data collected can be valuable, greater quality and breadth of relevant data such as patient conditions and outcomes are collected by medical insurers, which leads often to their prioritization as a source over private-sector facility and laboratory data.

Where medical insurers' data are unavailable or unattainable, facility and laboratory data can be considered as a substitute for similar application to risk prediction use cases.

d) Health care mobile applications data

A variety of health care worker and patient-facing mobile application platforms are available that enable direct communication with workers and patients and collect and manage data. Examples of these platforms include mobile applications for community health workers to track progress and services provided to patients in a community, patient referral applications linking primary health care facilities to specialists for easy referral processes, and messenger application platforms that enable direct communication with and feedback from patients, answering frequently asked questions and providing additional information and behavioral nudges through the platform.

These types of platforms collect a variety of patient health and health worker performance indicators and have clear adoption in emerging countries, an example of which is India. The data collected through these platforms has demonstrated potential to provide valuable insights for public health applications, such as studying the correlation between frequency and duration of community health worker visits and improvement in patient education and outcomes.

Health care mobile apps could be used when tailored data is needed for a use case. For example, these platforms could be used to collect data on household and patient demographic factors for a specific location to inform more precise risk prediction in the High-Risk Hotspotter, Fallout Forecaster, or Risk Rater, or enable case managers and clinicians to collect more

nuanced information on treatment adherence for the Decongestion Driver.

e) Epidemiological data

Information on local disease burden, such as prevalence and incidence and related covariates, can provide meaningful insights for a number of applications. The level of granularity that can be provided in disease burden data is an important and potentially game-changing factor for use cases, with high-value data sets able to provide data aggregated at small geographic radii.

For the top ten use cases, granular prevalence and incidence data can massively improve the accuracy of hotspot identification of at-risk populations for the High-Risk Hotspotter and the LTFU Locator.

Area-level data

a) Geospatial data

Many of the providers of socio-demographic data also provide geospatial information, including area boundaries on which data layers of socio-economic and behavioral variables are overlaid. Geospatial information in the form of aerial photographs, points of interest locations, coordinates of private- and public -sector facilities, etc., is also an important source of data for area-specific applications. One such example involves using facility location data and population and disease prevalence data to analyze the accessibility of health care services for the population in an area, considering factors such as average distance to the facility and potential overcrowding or underutilization of a facility. There are multiple providers of geospatial mapping application programming interfaces (APIs) available for subscription, but a narrower breadth of providers of accurate and comprehensive facility location information.

Geospatial data is critical to enable optimization of staffing across facilities in an area for the Staff Supporter and to provide insights into potential factors that contribute to risk of patients falling off treatment in the Fallout Forecaster and the LTFU Locator.

b) Traffic data

Traffic data is another aspect of location intelligence data similar to geospatial and weather that provides insights into area-level events or trends. Multiple traffic data providers exist in the market, largely operating on a subscription model to online platforms or APIs on which clients can gain access to historic or live traffic data. Information available includes road network information, traffic speed data, points of interest, and traffic dwell times. Vendors of traffic data often also provide geo-analytics services where data is overlaid with other

variables, such as income levels, to show, for example, which roads travel through higher-income neighborhoods and when traffic in those areas is at its peaks. This type of data is often used for route optimization by logistics companies and vehicle tracking by fleet managers and insurance companies and also in network optimization by chains such as banks or retail stores to identify branches that are underperforming in relation to opportunities presented by the level of activity in an area.

For our use cases, traffic data could provide alerts of disturbances in an area that could affect community health workers for the High-Risk Hotspotter and the LTFU Locator, as well as providing added insight into the potential activity in and around a facility at different times of the day, relevant for the Better Booker and the Staff Supporter.

c) Weather data

Historical and live data on key weather variables such as temperature, humidity, pressure, and cloud coverage are available from several identified sources, and similar to traffic data, these providers make data packages available for subscription through online platforms or APIs. Weather data, such as temperature and rainfall patterns, have been used in the past to predict and create early warning systems for tropical disease outbreaks.⁴

This type of data could be leveraged in the Staff Supporter and the Better Booker use cases to understand the impact that weather events such as heavy rains have on staff absenteeism and other performance indicators, as well as the associated impact of such events on number of patients in a facility. This data can also be used for community health worker alerts in the High-Risk Hotspotter and the LTFU Locator.

d) Crime data

The prevalence of crimes such as domestic violence and sexual abuse in an area is often related to elevated risks of HIV infection. In some countries, annual crime statistics across categories are publicly available and downloadable from police services websites. While these are an important source of data, some reports suggest that the rates of underreporting in crimes such as intimate partner violence and rape and sexual assault are high in this context—notably in LMICs—for reasons including perceived victim shame and fear of stigmatization. Therefore, consideration must be given to the limitations of reported crime statistics in such instances.

For the ten priority use cases detailed in this paper, crime data can provide insights into areas in which community health workers need to be provided better protection or in which to be more vigilant in the High-Risk Hotspotter

and the LTFU Locator. For these use cases, prevalence of certain crimes such as sexual violence can also inform the tailoring of intervention campaigns in hotspot areas to ensure the root causes of high HIV infections are adequately addressed. In the rare LMIC countries where more real-time crime report data is available (sometimes scraped from a Twitter, news, or website feed), this data can also be used to inform planned CHW outreach routes to optimize safety where possible.

e) Power outages data

Many LMICs experience power outages as growth in demand for electricity outpaces growth in supply. While some larger health care facilities are able to make contingencies, such as using backup generators in

the event of power cuts, smaller and/or less-resourced facilities are often less equipped to do so. These outages can have an effect on facility operations, such as delaying the update of electronic patient records requiring power or Internet connectivity and hindering patients' ability to arrive at scheduled appointments.

For our use cases, data on power outages can provide additional insight into drivers behind staffing attendance trends for the Staff Supporter, as well as insights into factors that result in missed appointments for the Better Booker use case. Further, this data can shed light onto potential reasons for discrepancies in reported treatment statistics for the Treatment Data Triangulator..

Footnotes

1. Fraym, "Machine learning to identify high-risk Covid-19 populations."
2. Facebook Covid-19 interactive map and dashboard.
3. Dean LT, et al., "Consumer credit, chronic disease and risk behaviours," *J Epidemiol Community Health* 73 (2019):73-78.
4. Hii YL, et al., "Forecast of dengue incidence using temperature and rainfall," *PLoS Negl Trop Dis* 6, no. 11 (2012): e1908, doi:10.1371/journal.pntd.0001908.

U.S. Agency for International Development

1300 Pennsylvania Avenue NW
Washington, DC 20523

Tel: (202) 712-0000

Fax: (202) 216-3524

www.usaid.gov