Improving HIV early infant diagnosis supply chains in sub-Saharan Africa: Models and application to Mozambique

Jérémie Gallien

Co-authors: Sarang Deo and Jónas Oddur Jónasson

In collaboration with NIH of Mozambique, and CHAI

London Business School

Senegal, November, 2015
The pediatric HIV epidemic

- 2 million children live with HIV
- 1800 children catch the virus via mother to child transmission (MTCT) every day
- Early treatment of HIV is available and effective, e.g. reducing mortality by 75%
- If untreated, half of MTCT infected infants die before the age of two

Chatterjee et al. 2011; Kourtis et al. 2006; Violari et al. 2008; Newell et al. 2004
HIV diagnosis in resource limited settings

Standard diagnostic method for adults

- Antibody blood tests at point-of-care clinics.

Diagnostic method for HIV exposed infants

- Dried blot spot (DBS) samples taken at point-of-care clinics.
- Polymerase Chain Reaction (PCR) analysis conducted at laboratories.

The resulting supply chain

Clinic Lab
HIV diagnosis in resource limited settings

**Standard diagnostic method for adults**
- Antibody blood tests at point-of-care clinics.

**Diagnostic method for HIV exposed infants**
- Dried blot spot (DBS) samples taken at point-of-care clinics.
- Polymerase Chain Reaction (PCR) analysis conducted at laboratories.

*The resulting supply chain*
An early infant diagnosis (EID) supply chain

An EID supply chain consists of:

- Clinics, where blood samples are drawn and results later communicated.
- Laboratories, where samples are processed.
- An assignment of clinics to labs.
  - A static partition in practice.
An early infant diagnosis (EID) supply chain

An EID supply chain consists of:

- Clinics, where blood samples are drawn and results later communicated.
- Laboratories, where samples are processed.
- An assignment of clinics to labs.
  - A static partition in practice.
An early infant diagnosis (EID) supply chain

An EID supply chain consists of:

- Clinics, where blood samples are drawn and results later communicated.
- Laboratories, where samples are processed.
- An assignment of clinics to labs.
  - A static partition in practice.
Main characteristics

Clinics

- Variability in arrivals of infants.
- Unpredictable transport opportunities from clinics.
- Variability in batch sizes.

Laboratories

- Random arrival of batches.
- Random size of arrival batches.
- Fixed size of processing batches.
- $\sum_{i=1}^{n} GI^{[X_i]} / G^{(b,b)} / c$ queue
Main characteristics

**Clinics**
- Variability in arrivals of infants.
- Unpredictable transport opportunities from clinics.
- Variability in batch sizes.

**Laboratories**
- Random arrival of batches.
- Random size of arrival batches.
- Fixed size of processing batches.
- \[ \sum_{i=1}^{n} GI[X_i] / G^{(b,b)} / c \text{ queue} \]
The importance of timing

Shortening result TATs affects health outcomes in three ways:

1. More infants can be reached before they die.
2. Treatment is more effective, both in terms of morbidity and mortality.
3. Caretakers are more likely to follow up for results for short TATs.

Newell et al. 2004; Violari et al. 2008; Latigo et al. 2013
The importance of timing

*Shortening result TATs affects health outcomes in three ways:*

1. More infants can be reached before they die.
2. Treatment is more effective, both in terms of morbidity and mortality.
3. Caretakers are more likely to follow up for results for short TATs.

Newell et al. 2004; Violari et al. 2008; Latigo et al. 2013
The importance of timing

*Shortening result TATs affects health outcomes in three ways:*

1. More infants can be reached before they die.
2. Treatment is more effective, both in terms of morbidity and mortality.
3. Caretakers are more likely to follow up for results for short TATs.

---

Newell et al. 2004; Violari et al. 2008; Latigo et al. 2013
## Research questions

### Relevant research areas for EID:

<table>
<thead>
<tr>
<th>Public health</th>
<th>Operations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medical: e.g. improved treatments</td>
<td>e.g. network structure</td>
</tr>
<tr>
<td>Technological: e.g. improved diagnostic equipment</td>
<td></td>
</tr>
<tr>
<td>Behavioral: e.g. to increase PMTCT participation</td>
<td></td>
</tr>
</tbody>
</table>

1. Impact of optimal assignment of clinics to labs?
   - TAT shortened by 11%, ART uptake increased by 4%

2. Impact of optimal allocation of diagnostic capacity?
   - TAT shortened by 22%, ART uptake increased by 7%
## Research questions

### Relevant research areas for EID:

<table>
<thead>
<tr>
<th>Public health</th>
<th>Medical: e.g. improved treatments</th>
<th>Technological: e.g. improved diagnostic equipment</th>
<th>Behavioral: e.g. to increase PMTCT participation</th>
</tr>
</thead>
</table>

### 1. Impact of optimal assignment of clinics to labs?
- TAT shortened by 11%, ART uptake increased by 4%

### 2. Impact of optimal allocation of diagnostic capacity?
- TAT shortened by 22%, ART uptake increased by 7%
Research questions

Relevant research areas for EID:

<table>
<thead>
<tr>
<th>Public health</th>
<th>Operations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medical: e.g. improved treatments</td>
<td>e.g. network structure</td>
</tr>
<tr>
<td>Technological: e.g. improved diagnostic equipment</td>
<td></td>
</tr>
<tr>
<td>Behavioral: e.g. to increase PMTCT participation</td>
<td></td>
</tr>
</tbody>
</table>

1. Impact of optimal assignment of clinics to labs?
   - TAT shortened by 11%, ART uptake increased by 4%

2. Impact of optimal allocation of diagnostic capacity?
   - TAT shortened by 22%, ART uptake increased by 7%
**OM literature**

*The classical facility location problem*
- Daskin (2008)

*The facility location problem to maximize participation*

*The congested facility location problem*
- Zhang et al. (2009), Marianov et al. (2008), Brimberg et al. (1997), Brimberg and Mehrez (1997)

*Here*
- Congested facility location problem with batching in arrivals and processing.
Overview of EID practices
- Creek et al. (2007), Chatterjee et al. (2011)

Issues with participation in EID systems
- Ciaranello et al. (2011a), Mmbaga et al. (2009), Msuya et al. (2006), Latigo et al. (2013)

Epidemiological studies
- Ciaranello et al. (2011b), Creese et al. (2002), Aledort et al. (2006), Stevens et al. (2008), Menzies et al. (2009)

Estimation of mortality in infected infants
- Marston et al. (2005), Newell et al. (2004)

Here
- Quantitative analysis of the effect of operations on public health.
Overview of analysis

*Simulation model*

- Descriptive of real EID system.

*Optimization model*

- Prescriptive for operational interventions.

*Mozambique data*

- Input data for both models.
- Output data
  - to validate prediction accuracy of simulation.
  - to evaluate performance impact of optimization.
Simulation model details

Main performance metrics

- The results turnaround times (TATs).
- The percentage of infected infants initiating treatment.

Scope

- The system is simulated at the country level.
Simulation model details

*The network*

- Set of clinics
- Set of laboratories
- Assignment of clinics to labs
- Travel time
Simulation model details

**The infants**

- Age at testing
- PMTCT regimen
- HIV status (depending on PMTCT status)
- Age at infection (if infected)
- Life expectancy of infant if untreated (Newell et al. 2004)
- Will caretaker follow up for the results (Latigo et al. 2010)
- Initiation of infant on treatment (if infected and alive) (Ciaranello et al. 2011)
Simulation model details

**The clinics (pre-lab)**

- Arrival rate of infants (*Zero-Inflated Poisson*)
- Arrival rate of transport opportunities (*Poisson*)

---

<table>
<thead>
<tr>
<th>Birth</th>
<th>Blood drawn</th>
<th>Sample dispatched</th>
<th>Sample received</th>
<th>Lab batch ready</th>
<th>Server available</th>
<th>Processing finished</th>
<th>Results dispatched</th>
<th>Results received</th>
<th>Results ready</th>
<th>Results communicated</th>
<th>Treatment starts</th>
</tr>
</thead>
</table>

---

Sample turnaround time

Communication turnaround time
Simulation model details

*The laboratories*

- Size of processing batch
- Number of servers
- Service time distribution (*Erlang*)
- Post-processing delay at the laboratory

Hence, each laboratory can be described as a \( \sum_{i=1}^{n} GI[X_i] / G(b,b) / c \) queue
Simulation model details

The clinics (post-lab)

- Post-processing delay at the clinic
- Delay from receipt of results until initiation of treatment
The caretaker behavior
(Deo et al. 2014, Latigo et al., 2014)

TAT impact:

- Results in 1st month: Follow up = 61%
- Results in 2nd month: Follow up = 58%
- Results in over 2 months: Follow up = 51%

PMTCT impact:

- PMTCT participants 81% more likely to follow up than non participants.
Setting

**Mozambique**
(www.unaids.org)

- HIV prevalence: 11.3%
- People living with AIDS: 1.4m
- Annual new infections: 130,000
- Annual AIDS deaths: 74,000

**EID Network**

- 4 Labs
- 410 Clinics
- 33,322 tests in 2011
- 4,007 positive samples
- About 39% follow up for results

**Collaborators**

- National Institute of Health (NIH) and Clinton Health Access Initiative (CHAI)
Validation of prediction accuracy

*Simulation model is validated in 4 main ways:*

1. Mean TATs and LCTs
2. Distribution of TATs
3. Transportation times
4. Fraction of infected infants treated

} Operations

} Public health
Validation of prediction accuracy

*Simulation model is validated in 4 main ways:*

1. Mean TATs and LCTs
2. Distribution of TATs
3. Transportation times
4. Fraction of infected infants treated

### Operations

<table>
<thead>
<tr>
<th></th>
<th>Maputo</th>
<th>Nampula</th>
<th>Quelimane</th>
<th>Beira</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinic delay</td>
<td>5</td>
<td>10</td>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>Transportation times</td>
<td>20</td>
<td>25</td>
<td>30</td>
<td>20</td>
</tr>
<tr>
<td>Lab Cycle Times</td>
<td>25</td>
<td>30</td>
<td>35</td>
<td>30</td>
</tr>
</tbody>
</table>

### Public health

<table>
<thead>
<tr>
<th></th>
<th>Maputo</th>
<th>Nampula</th>
<th>Quelimane</th>
<th>Beira</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinic delay</td>
<td>8</td>
<td>12</td>
<td>16</td>
<td>8</td>
</tr>
<tr>
<td>Transportation times</td>
<td>25</td>
<td>30</td>
<td>35</td>
<td>30</td>
</tr>
<tr>
<td>Lab Cycle Times</td>
<td>30</td>
<td>35</td>
<td>40</td>
<td>40</td>
</tr>
</tbody>
</table>
Validation of prediction accuracy

Simulation model is validated in 4 main ways:

1. Mean TATs and LCTs
2. Distribution of TATs
3. Transportation times
4. Fraction of infected infants treated

Operations

Public health

<table>
<thead>
<tr>
<th>Result TAT</th>
<th>Field data</th>
<th>Simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 month</td>
<td>40%</td>
<td>40%</td>
</tr>
<tr>
<td>2 months</td>
<td>30%</td>
<td>30%</td>
</tr>
<tr>
<td>3 months or more</td>
<td>10%</td>
<td>10%</td>
</tr>
</tbody>
</table>
Validation of prediction accuracy

*Simulation model is validated in 4 main ways:*

1. Mean TATs and LCTs
2. Distribution of TATs
3. Transportation times
4. Fraction of infected infants treated

}{ Operations
}{ Public health
Validation of prediction accuracy

*Simulation model is validated in 4 main ways:*

1. Mean TATs and LCTs
2. Distribution of TATs
3. Transportation times
4. Fraction of infected infants treated

{ Operations  
Public health

---

**Uptake of ART**

<table>
<thead>
<tr>
<th>Proportion of infants starting ART</th>
<th>3 PCT Pilot Cohort Data</th>
<th>Simulation model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>38.9%</td>
<td>40.9%</td>
</tr>
</tbody>
</table>
The EID network design problem

Max Treatment initiation fraction(sample TATs)

where: sample TAT = exogenous waits + travel times + lab cycle times

lab cycle times(lab utilization)

lab utilization(assignment)

travel times(assignment)
The intractable EID network design problem

maximize \( \nu \sum_{i=1}^{l} \pi_i \lambda_i \left( \sum_{k \in \{P, NP\}} Q_{i,k} V_k E \left[ F_k (\zeta, TAT_i, \eta, \delta) \right] \right) \)

subject to:

\[ \sum_{j=1}^{J} z_{ij} = 1 \quad \forall i \in I \]  
\{ Assignment \}

\[ \rho_j = \frac{\sum_{i=1}^{l} z_{ij} \pi_i \lambda_i E [S_j]}{b_j c_j} \quad \forall j \in J \]  
\{ Utilization \}

\[ \rho_j \leq 1 - \epsilon \quad \forall j \in J \]  
\{ Lab sojourn time \}

\[ L_j = B_j (\rho_j) + W_j (\rho_j) + P_j \quad \forall j \in J \]  
\{ Lab sojourn time \}

\[ TAT_i = C_{i, pre} + \sum_{j=1}^{J} z_{ij} T_{ij} + \sum_{j=1}^{J} z_{ij} L_j + C_{i, post} \quad \forall i \in I \]  
\{ Turnaround time \}

\[ z_{ij} \in \{0, 1\}; \rho_j \in \mathbb{R}_+ \]  
\{ Decision variables \}
Reformulation: 2. Waiting time in $\sum_{i=1}^{n} GI[X_i] / G(b,b) / c$ queues

Current constraint

$$L_j = B_j(\rho_j) + W_j(\rho_j) + P_j \quad \forall j \in J$$


$$E[W](\rho_j) = \frac{(b_j - 1)E[S_j]}{2b_j c_j \rho_j} + \frac{E[S_j]}{b_j c_j} \cdot \frac{\rho_j \sqrt{2(c_j+1) - 1}}{1 - \rho_j} \left( B + \frac{1}{2} + \frac{b_j SCV[S_j]}{2} \right) + E[S_j]$$

- Batch – formation delay
- Congestion delay
- Processing time
Reformulation: 2. Waiting time in $\sum_{i=1}^{n} GI[X_i] / G^{(b,b)} / c$ queues

Current constraint

$$L_j = B_j(\rho_j) + W_j(\rho_j) + P_j \quad \forall j \in J$$


$$E[W](\rho_j) = \frac{(b_j - 1)E[S_j]}{2b_j c_j \rho_j} + \frac{E[S_j]}{b_j c_j} \cdot \frac{\rho_j^{\sqrt{2(c_j+1)}-1}}{1 - \rho_j} \left( B + \frac{1}{2} + \frac{b_j SCV[S_j]}{2} \right) + E[S_j]$$

Waiting time approximation
Reformulation: 2. Waiting time in $\sum_{i=1}^{n} GI[X_i] / G^{(b,b)} / c$ queues

Current constraint

$$L_j = B_j(\rho_j) + W_j(\rho_j) + P_j \quad \forall j \in \mathcal{J}$$


$$E[W](\rho_j) = \frac{(b_j - 1)E[S_j]}{2b_j c_j \rho_j} + \frac{E[S_j]}{b_j c_j} \cdot \frac{\rho_j}{1 - \rho_j} \left( B + \frac{1}{2} + \frac{b_j SCV[S_j]}{2} \right) + E[S_j]$$

**Batch – formation delay**

**Congestion delay**

**Processing time**

Linear waiting time approximation
Reformulation: 2. Waiting time in $\sum_{i=1}^{n} GI[X_i] / G^{(b,b)} / c$ queues

Current constraint

$$L_j = B_j(\rho_j) + W_j(\rho_j) + P_j \quad \forall j \in \mathcal{J}$$


$$E[W](\rho_j) = \frac{(b_j - 1)E[S_j]}{2b_j c_j \rho_j} + \frac{E[S_j] \rho_j \sqrt{2(c_j + 1) - 1}}{b_j c_j (1 - \rho_j)} \left( B + \frac{1}{2} + \frac{b_j SCV[S_j]}{2} \right) + E[S_j]$$

New constraints

$$\sum_{d=1}^{D} \tau_{jd} d \geq \alpha_{ju} \rho_j + \beta_{ju} \quad \forall j \in \mathcal{J}, u \in \mathcal{U}$$

$$\sum_{d=1}^{D} \tau_{jd} = 1 \quad \forall j \in \mathcal{J}$$
maximize \( \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{d=1}^{D} z_{ijd} \left( \sum_{m=1}^{M} \phi_{dm} \omega_{im} \right) \)

subject to:

\[ \sum_{j=1}^{J} \sum_{d=1}^{D} z_{ijd} = 1 \quad \forall i \in I \]

\[ \rho_j = \frac{\sum_{i=1}^{I} \sum_{d=1}^{D} z_{ijd} \pi_i \lambda_i E [S_j]}{b_j c_j} \quad \forall j \in J \]

\[ \rho_j \leq 1 - \epsilon \quad j \in J \]

\[ \sum_{d=1}^{D} \tau_{jd} d \geq \alpha_{ju} \rho_j + \beta_{ju} \quad \forall j \in J, u \in U \]

\[ \sum_{d=1}^{D} \tau_{jd} = 1 \quad \forall j \in J \]

\[ z_{ijd} \leq 0 \quad \forall i \in I, j \in J, 0 \leq d \leq \Delta_{ij} \]

\[ z_{ijd} \leq \tau_{j(d-\Delta_{ij})} \quad \forall i \in I, j \in J, \Delta_{ij} < d \leq D \]

\[ z_{ijd}, \tau_{jd} \in \{0, 1\}; \rho_j \in \mathbb{R}_+ \]

*See paper for the Optimal Capacity Allocation extension and computational cuts.*
Main numerical results

Operational configurations

- **Status quo**
- **Mild intervention:** Optimal Lab Assignment (OLA)
- **Radical intervention:** Optimal Capacity allocation (OCA)

![Treatment initiation graph](image)

![Turnaround time graph](image)
Main numerical results

**Operational configurations**

- **Status quo**
- **Mild intervention**: Optimal Lab Assignment (OLA)
- **Radical intervention**: Optimal Capacity allocation (OCA)

![Graph showing treatment initiation and turnaround time by operational configuration](image)

**Treatment initiation**

- Status quo: 40%
- OLA: 41%
- OCA: 42%

**Turnaround time**

- Status quo: 30 days
- OLA: 29 days
- OCA: 28 days

Proportion of treatment initiation across operational configurations:

- Status quo: 41%
- OLA: 42%
- OCA: 43%

Turnaround time decrease:

- Status quo to OLA: 11% decrease

Main numerical results

**Operational configurations**

- **Status quo**
- **Mild intervention**: Optimal Lab Assignment (OLA)
- **Radical intervention**: Optimal Capacity allocation (OCA)

![Chart showing treatment initiation and turnaround time](chart.png)
Optimization for administrative districts and regions

Clinic level
- Improvement: 4%

District level
- Improvement: 4%

Region level
- Improvement: 2.5%
Further insights for capacity allocation

*Utilization sensitivity*

- Vary arrival rates to scale utilization from 5% to 95%.

**Improvement over status quo:**

*Infants initiating treatment*

![Graph showing improvement over status quo for infants initiating treatment.](image)
Further insights for capacity allocation

Transportation time sensitivity

• Vary transportation times from a network average of 3 to 46.

![Graph showing transportation time sensitivity with three different lines representing OCA, Con[N], and Status quo. The x-axis represents the average network transportation time, and the y-axis represents the percentage of infected infants treated. The graph includes a note indicating the average transportation time in Mozambique.]
Conclusions and next steps

Conclusion

• We present a detailed simulation model to test various interventions in EID systems.
• We develop a new approximation for the wait time in a $\sum_{i=1}^{n} GI[X_i] / G^{(b,b)} / c$ queue.
• We develop an analytical model for solving the EID network facility location problem.
• We evaluate the public health impact of the operational structure of the EID network.

Answers to research questions

1. Impact of optimal assignment of clinics to labs?
   • TAT shortened by 11%, ART uptake increased by 4%
2. Impact of optimal allocation of diagnostic capacity?
   • TAT shortened by 22%, ART uptake increased by 7%
Conclusions and next steps

Conclusion

• We present a detailed simulation model to test various interventions in EID systems.
• We develop a new approximation for the wait time in a \( \sum_{i=1}^{n} G_{I[X_i]} \| G^{(b,b)}/c \) queue.
• We develop an analytical model for solving the EID network facility location problem.
• We evaluate the public health impact of the operational structure of the EID network.

Next steps

• Health care delivery in resource limited settings
  • Detailed models
  • Non-trivial operational questions
  • Link to public health
• E.g. Community Health Workers
THANK YOU