## NHLS Data Strategy

Professor Wendy Stevens

Department of Molecular Medicine and Haematology

University of the Witwatersrand, National Priority Program,

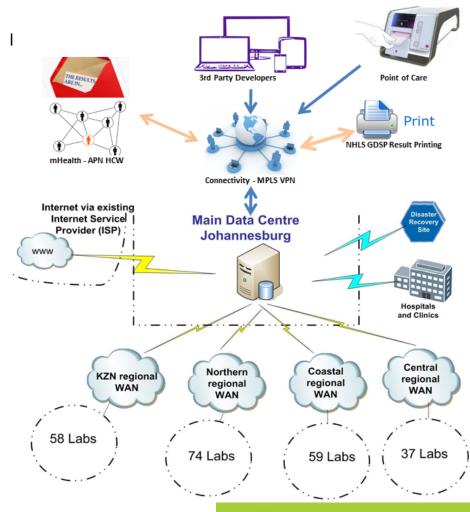
NHLS, Johannesburg, south Africa





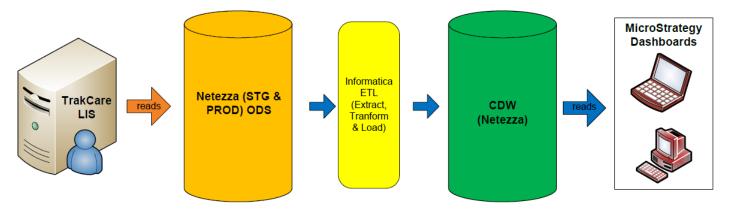
## Introduction

- NHLS has a unique repository of laboratory information
  - The data is representative of pathology information
     for +/- 80% of the South African population
- 270+ sites across South Africa
- Single Laboratory System
- Single Data Warehouse





## Laboratory Information Systems



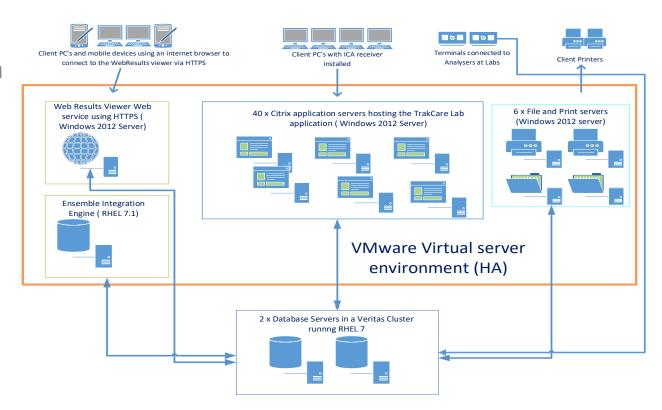
- TrakCare Lab
  - used in all labs in NHLS, NICD and NIOH
- Country-wide since Oct 2015
- Technical
  - Application hosted on Citrix Application servers
  - Supported by Caché Database
  - Uses Ensemble Integration engine
- Network
  - To all sites
  - MPLS implementation
  - SITA CAP connection

- All instruments are interfaced through a LIS (4 regional centers and connected through wide area network)
- CDW raw data is aggregated
- A disaster recovery site is based within 20km.
- Hospital information systems (HIS) use HL7 protocol and interface to the data center.
- All patient laboratory results are made available to HCW via an internet service provider (ISP) such as
- Virtual private networks (MPLS VPN) and other network traffic are routed for mHealth, point of care, SMS printer devices, etc



## Information held in CDW

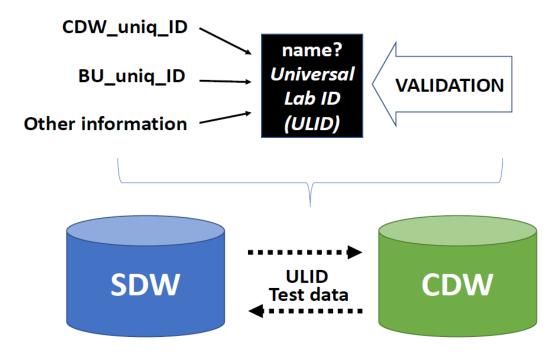
- Laboratory information:
  - Disa\*Lab legacy laboratory information system
  - TrakCare Lab new laboratory information system
  - Information from private laboratories
- Business Information
  - Financial information
  - Lab staffing information
- National statistics
  - Population statistics
  - Prevalence surveys (e.g. antenatal HIV)
  - DHIS information





## Surveillance and Corporate data warehouse

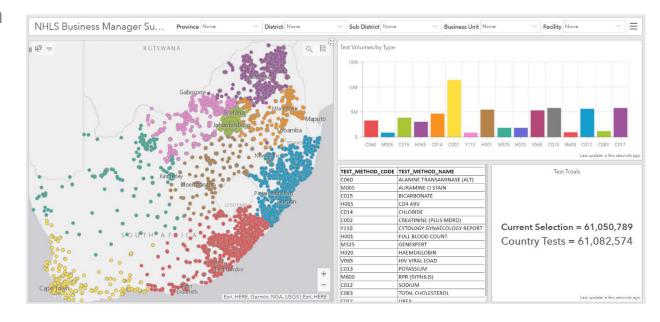
In future: Algorithm integration into SDW and CDW





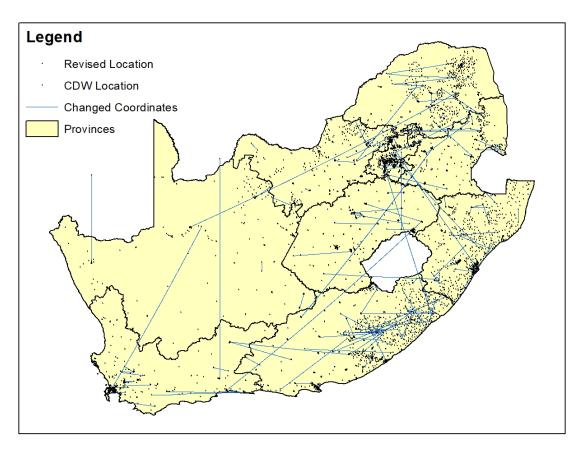
## **Data Capabilities**

- Process +/- 300 million raw data records per month
- Perform extensive data cleansing and transformation
- Record linking Laboratory information systems are specimen-centric and therefore need to build a patient or case view of the data
  - Implemented data linking program
  - Uses fuzzy logic to evaluate demographic attributes and assign a probabilistic % match
  - Automatically allocates a patient identifier to all matches scoring above a given threshold
- Reporting analytical, formatted, integrated spatial
   & dashboard reporting is available





## **Facility coordinate cleaning**



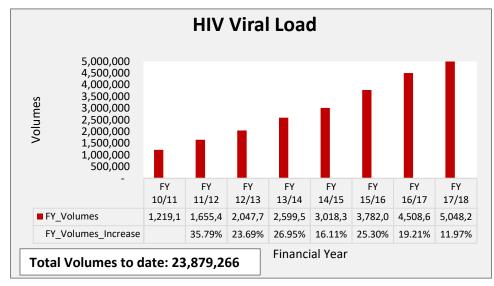
#### Process in constant Evolution

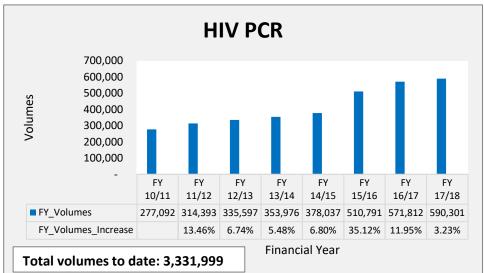
- **Simple exercise of mapping**: clinics, laboratories, drive time, volumes, results, TAT, analyzer volumes
- Service Mapping Network: logistics: route selection, physical location, GIS
- Mapping all this into an LIS/LIMS
- Mapping instruments, tests and test codes to LIS
- Mapping demographics and result to location (LIMS)
- Added layer of Sophistication
- Mapping to a Central Data Warehouse (CDW)
- "Geospatial mapping": fixed and now mobile

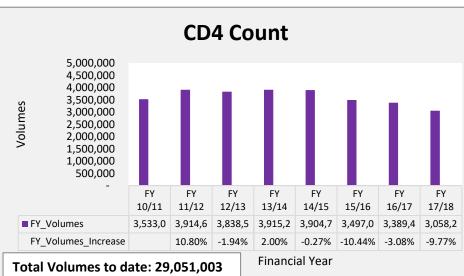


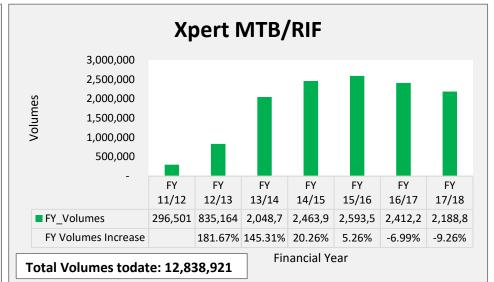
## Across the National Priority Program

## 80% population, networked, instruments interfaced -1 LIS

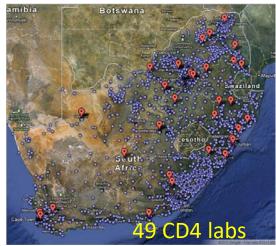










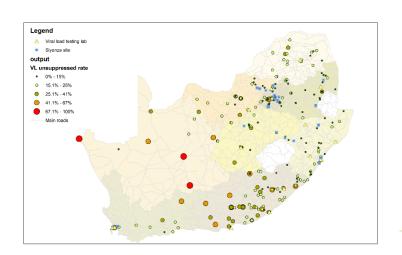


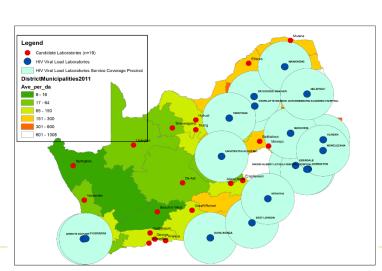




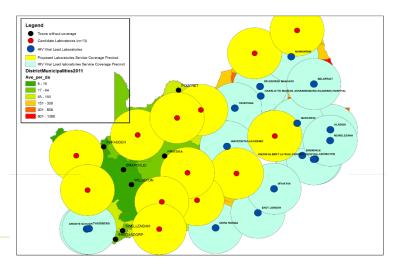
## Data applications beyond result reporting

- Operational dashboards
- Surveillance
- Gaps in services (coverage)
- GIS mapping for route optimization
- Placement of instruments within the tiered laboratory framework
- Placement of POC
- HIV Longitudinal cohort development (program success)
- Integrate HIV/TB cohort
- Molecular granularity for TB control

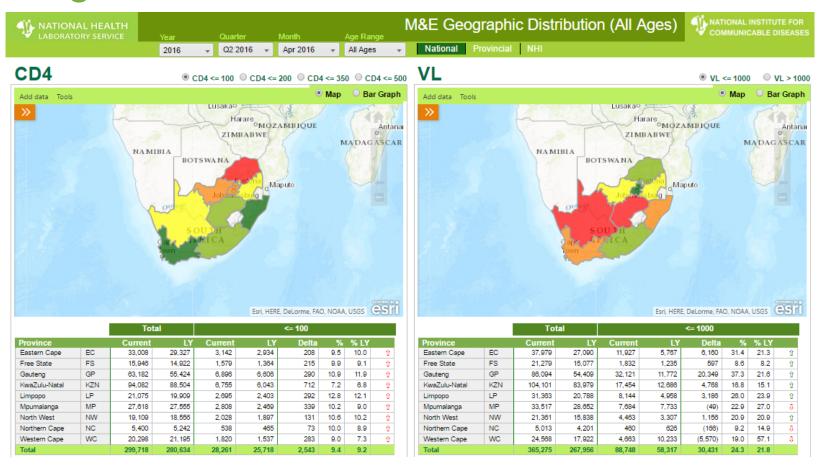








## Program value











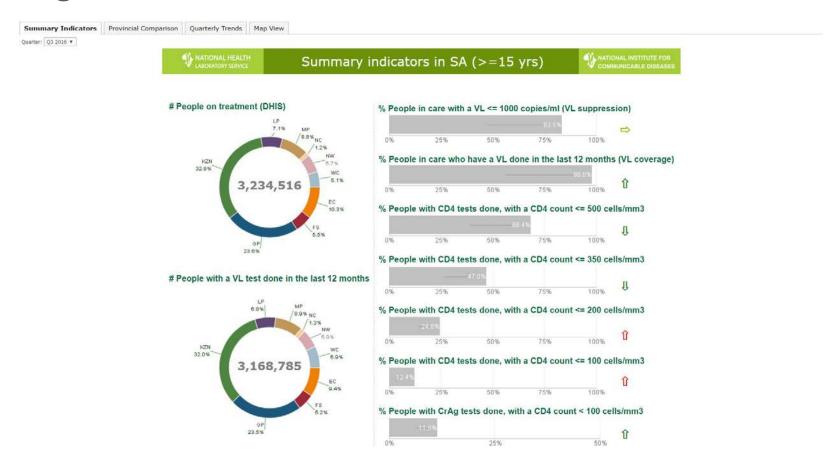








## Program value



















## Building the NHLS National HIV Cohort

#### INPUT · Lab episodes, with identifying demographics

#### 1. Pre-process data

- Cleaning Standardization
- · Reduction to exact matches on first/last/DOB/sex/facility

#### 2. Search for edges

- · Exact match on inversions,
- multiple names, nicknames · Fuzzy matching within blocks to reduce comparisons

#### 3. Score edges

- Jaro-Winkler string
- comparisons for names
- · Fellegi-Sunter similarity scores
- · Optimized weights

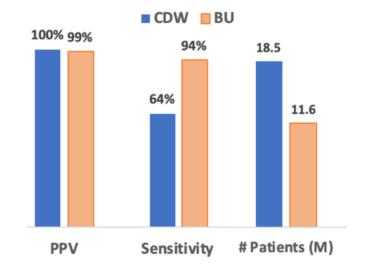
#### 4. Link and resolve entities

- · Thresholds for matches
- Transitivity
- · Graph-based techniques

#### OUTPUT

- Unique Patient Identifier (BU\_uniq\_ID)
- · Cluster characteristics for sensitivity analysis

National HIV Cohort	
# Lab Tests	98 M
# Specimens	71 M
# Patients	11.6 M
# on ART	4.2 M



#### **Building a National HIV Cohort from Routine Laboratory Data: Probabilistic Record-Linkage with Graphs**



Jacob Bor, William MacLeod, Katia Oleinik, James Potter, Alana T. Brennan, Sue Candy, Mhairi Maskew, Matthew P. Fox, Ian Sanne, Wendy S. Stevens, Sergio Carmona

doi: https://doi.org/10.1101/450304



## Research using the cohort (example 1)

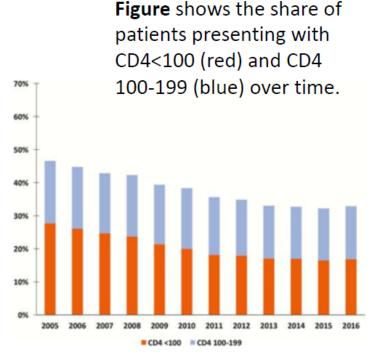
### Clinical Infectious Diseases

Persistent High Burden of Advanced HIV Disease Among Patients Seeking Care in South Africa's National HIV Program: Data From a Nationwide Laboratory Cohort &

Sergio Carmona ™, Jacob Bor, Cornelius Nattey, Brendan Maughan-Brown,
Mhairi Maskew, Matthew P Fox, Deborah K Glencross, Nathan Ford, William B MacLeod

Clinical Infectious Diseases, Volume 66, Issue suppl\_2, 1 April 2018, Pages S111–S117, https://doi.org/10.1093/cid/ciy045

Published: 04 March 2018





## Research using the cohort (example 2)

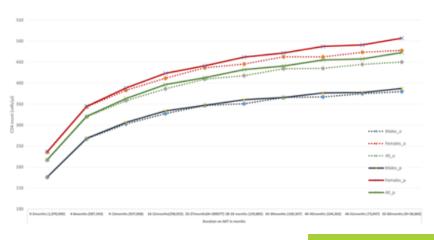


RESEARCH ARTICLE

CD4 count recovery and associated factors among individuals enrolled in the South African antiretroviral therapy programme: An analysis of national laboratory based data

Tendesayi Kufa<sup>6</sup>, Zara Shubber<sup>3</sup>, William MacLeod<sup>5</sup>, Simbarashe Takuva<sup>6</sup>, Sergio Carmona<sup>7</sup>, Jacob Bor<sup>4</sup>, Marelize Gorgens<sup>3</sup>, Yogan Pillay<sup>8</sup>, Adrian Puren<sup>1,9</sup>, Jeffrey W. Eaton<sup>10</sup>, Nicole Fraser-Hurt<sup>3</sup>

Figure shows CD4 trajectories among men (bottom), women (top), and both (middle) on ART





## Research using the cohort (example 3)

## Analysis of Big Data for better targeting of ART Adherence Strategies

Spatial clustering analysis of viral load suppression by South African province, district, sub-district and facility (April 2014 – March 2015)

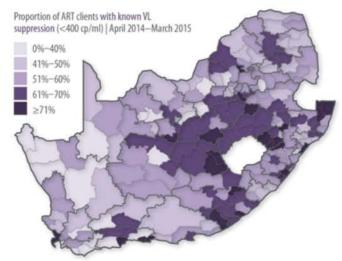
November 2015





William MacLeod, Jacob Bor, Kathryn Crawford, and Sergio Carmona with NDOH and World Bank collaborators

#### Viral suppression results: Identifying success Can we learn from the-dark shaded sub-districts?





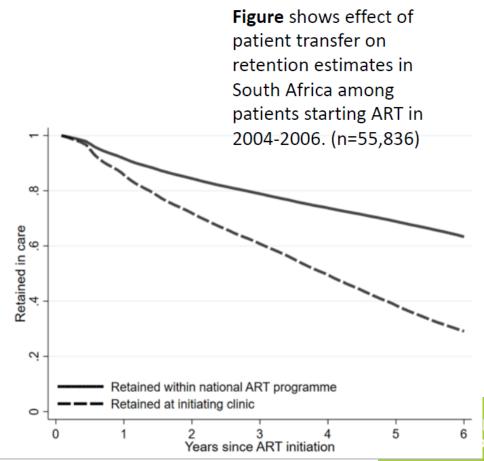
## Research using the cohort (example 4)



RESEARCH ARTICLE

Estimating retention in HIV care accounting for patient transfers: A national laboratory cohort study in South Africa

Matthew P. Fox<sup>1,2,3</sup>\*, Jacob Bor<sup>2</sup>, Alana T. Brennan<sup>2,3</sup>, William B. MacLeod<sup>2,3</sup>, Mhairi Maskew<sup>3</sup>, Wendy S. Stevens<sup>4,5</sup>, Sergio Carmona<sup>4</sup>



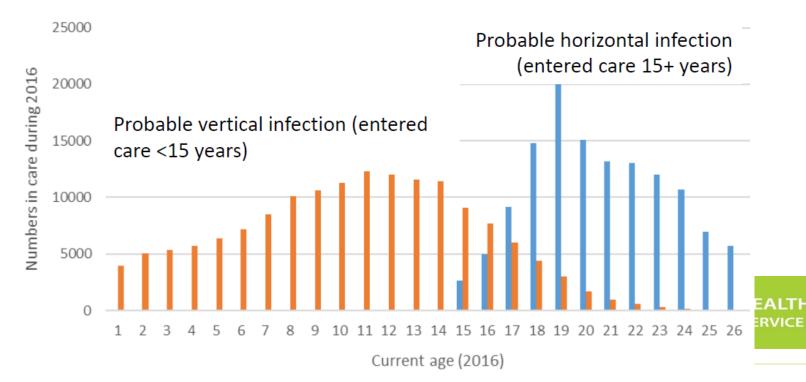
## Research using the cohort (example 5)

The Adolescent HIV Treatment Bulge in South Africa's National HIV Program: a Retrospective National Cohort Study

Mhairi Maskew, Jacob Bor, William MacLeod, Sergio Carmona, Gayle G. Sherman, Matthew P. Fox

In press

THE LANCET HIV



## Extensions of the HIV cohort linkage

- 1. How does the algorithm perform for other conditions?
  - Ongoing work with the NPP and NICD TB groups to link and validate for TB (BU/WITS R01 under review)
  - Recently funded work with Jaya George and Nigel Crowther to look at NCDs (Alana Brennan K01)
  - Research opportunities to look at HIV/TB/NCD multi-morbidity
- 2. Can performance be improved for infant and paediatric cases?
  - Ongoing work with Gayle Sherman
- 3. Can the algorithm be integrated into real-time linkage in the data warehouse, and how does it perform when scaled to all conditions?
  - Ongoing work with NICD
  - For pragmatic reasons, pilot at SDW, then move to CDW if successful



## Potential future opportunities

- Can NHLS National HIV Cohort be linked to TIER and other data sources to build integrated cohort to improve surveillance/research?
  - We have proposed some pilot work
- Can algorithm be leveraged to improve delivery of patient care?

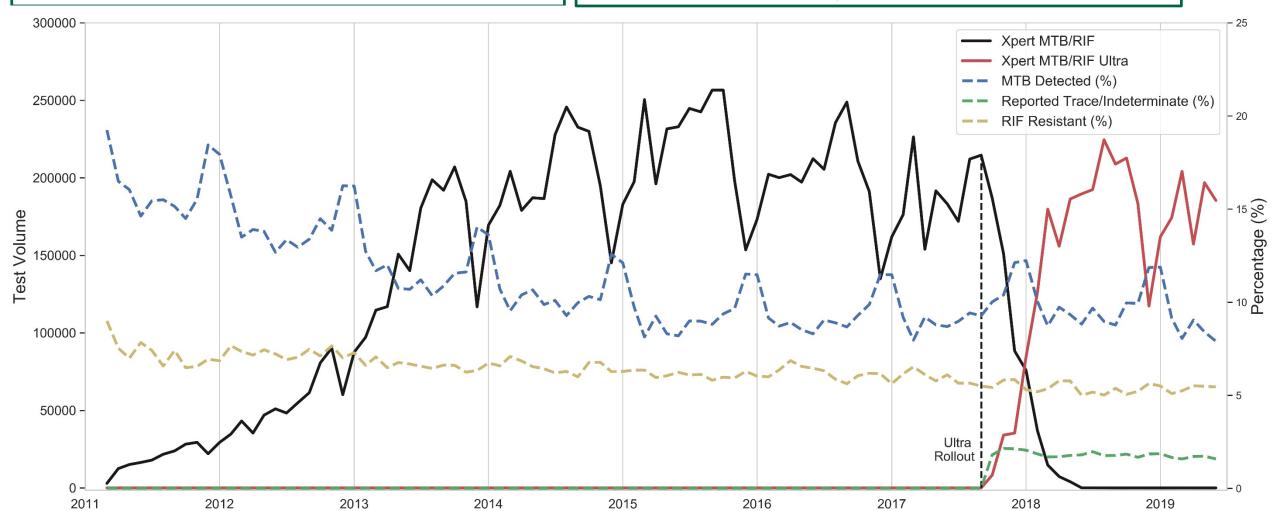


## Lots of Potential for Collaborative Research

- Here are some focus areas: (Are these the right ones? Are there others?)
- 1. VL monitoring and treatment outcomes
- 2. Paediatric HIV and pregnancy cohort
- 3. HIV drug resistance
- 4. HIV advanced disease
- 5. TB and HIV/TB co-infection
- 6. Non-communicable diseases
- 7. Data linkages and technical

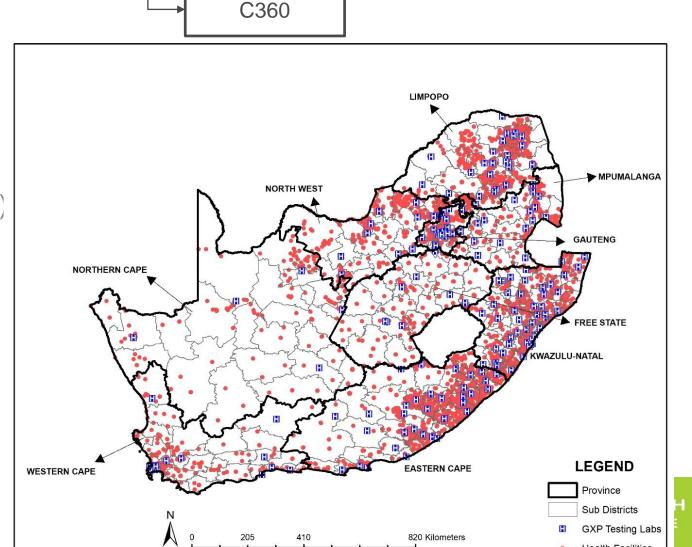


- CDW (March 2011 June 2019)
  - > 12.3 million Xpert MTB/RIF tests
  - > 3.2 million Xpert MTB/RIF Ultra tests
- C360 (June 2016 July 2019)
  - >1.8 million Xpert MTB/RIF tests
  - >2.4 million Xpert MTB/RIF Ultra Tests



## **Data Architecture**

- Data is transferred in two pathways:
  - TrakCare, Laboratory
     Information System (LIS),
     Central Data Warehouse (CDW)
  - 2. C360 Cepheid's platform.
- The data from CDW is comprehensive
- The data from C360 is at the laboratory level



**CDW** 

LIS

GeneXpert Device

#### **CDW DATA**

Lab no Referring lab no Testing lab name **Testing lab province** Province **Health district** Health sub district **Facility name Printed res** Specimen type Rif res MTB Raw result PROBE A Ct value PROBE B Ct value PROBE A Ct value PROBE B Ct value PROBE C Ct value PROBE D Ct value PROBE E Ct value SPC Ct value **Tested age years Tested date** Month no Year Registered date Reviewed date TAT Inlab DCS facility name **DCS Location name DCS Provice** Result Rif MTB Result Description CDW

Central data warehouse

Size of circles indicative of

data size and the data can

be further categorised into

instrument, assay, user

groups

**C360** 

Supplier Dashboard C360 DATA

System name
System serial number
System Model
Institution
Assay name
Assay version
Assay group

Result code

Result code

Status GUID

Reagent lot ID

Expiration date Cartridge S/N

Test type Start time

End time
Export time
Sample type

System user login System user name Software version

Ambient temp
Error status

Main error code
All error codes

All errors Lab identifier Lab name

Lab address
Lab city
Lab state
Lab country

Lab region
Lab postal code
Location country
Location region

Location district
Location subdistrict

Timezone

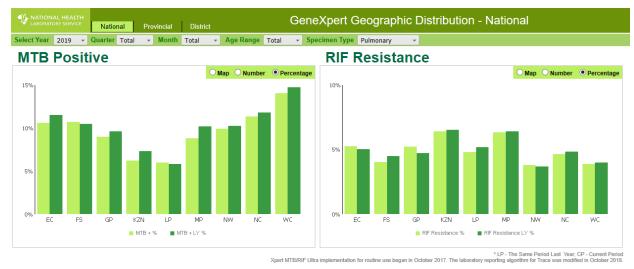
#### **OVERLAP DATA**

Testing lab name
Testing lab province
Health district
Health sub district
Rif res
PROBE A Ct value
PROBE B Ct value

PROBE A Ct value
PROBE B Ct value
PROBE C Ct value
PROBE D Ct value
PROBE E Ct value
SPC Ct value
Tested date
Result Rif
MTB Result Description

## How is the data used and potential future

TB Dashboards for the NDoH

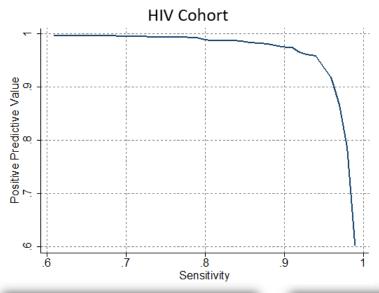


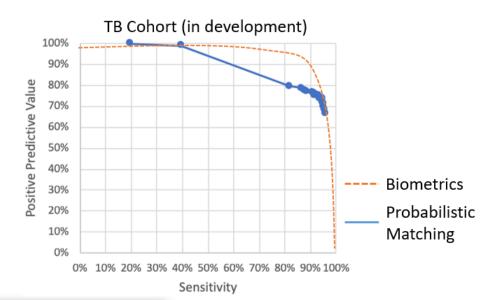
5,977 9,753 286 2.4% 0.3% 56 001 187,182 28,269 441,758 5,582 1,234 1.3% 0.2% 9,476 44 239 7.3% 2,052 95.734 9,087 6.0% 5.8% 934 190 1.0% 0.1% 0 1,512 0.0% 1.0% 3,071 2.1% 468 4.8% 66,873 1,605 433 2.4% 0.4% 0 1,539 1,959 2,004 71,007 7,025 12,245 1,460 287 2.1% 0.2% 0 2,051 0.0% 1.7% 2,229 2,837 3.1% 2.4% 449 3.8% 765 677 877 1.6% 1.4% 222 361 4.6% 4.8% 63 914 7 505 162 1.8% 0.3% 0 987 0.0% 1.5% 756 2.2% 0.3% 0 4,255 0.0% 1.8% 2,710 4,053 1.7%



## In the absence of a unique ID?

## The (Potential) Promise of Biometrics





skiftet prepriet find possied ordine Nov. 2, 01th, doi: http://dx.doi.org/10.1501/40004. The copyright holder for this prepriet (which was not peer-reviewed) is the sufficient new has preside blankly a fixment to fishing the prepriet in peopletally. It is made evaluable under a CC-8Y-NC-HD 4.0 International foreign.

Working Paper. Correspondence to bor@bu.edu. Version: October 10, 2018

Building a National HIV Cohort from Routine Laboratory Data: Probabilistic Record-Linkage with Graphs

Jacob Bor<sup>1,2,3-5</sup>, William MacLeod<sup>1,2-</sup>, Katia Oleinik<sup>4</sup>, James Potter<sup>4</sup>, Alana T. Brennan<sup>1,2</sup>, Sue Candy<sup>4</sup>, Mhairi Maskew<sup>2</sup>, Matthew P. Fox<sup>1,2,3</sup>, Ian Sanne<sup>6</sup>, Wendy S. Stevens<sup>5</sup>, Sergio Carmona<sup>5</sup>

<sup>1</sup> Department of Global Health, Boston University School of Public Health, USA, <sup>3</sup> health Economics and Epidemiology Research Office, Department of Internal Medicine, School of Clinical Medicine, Faculty of Health Sciences, University of Witwasters and, South Africa, <sup>3</sup> Department of Epidemiology, Boston University School of Public Health, USA, <sup>5</sup> Research, Computing Service, Boston University, USA, <sup>3</sup> Resolute Health Laboratory Service, South Africa, and Department of Molecular Medicine and Haematology, University of the Witwastersrand, <sup>5</sup> Rigist to Care, South Africa,

[qual contributions. 1 jbor@bu.edu

# PESSANCHARTICLE Estimating retention in HIV care accounting for patient transfers: A national laboratory cohort study in South Africa Matthew P. Fox \*3.3\*\*, Jacob Bor\*, Alana T. Brennan\*3, William B. MacLeod\*3, Mhairi Masker\*, Wendy S. Bevens\*1, Sergio Carmona\* 1 Department of Epidemiology, Boston University School of Public Health, Boston, Massachusetts, United States of America, 3 Health Economics and Epidemiology Research Office, Department of Health Economics of Chinal Health, Boston, Massachusetts, United States of America, 3 Health Economics and Epidemiology Research Office, Department of Michocal Medicine and Haematology, University of the Witnessersand, Johannessersand, South Micros. John Micros. South Micros. John Micros. South Micros. John Micros. J

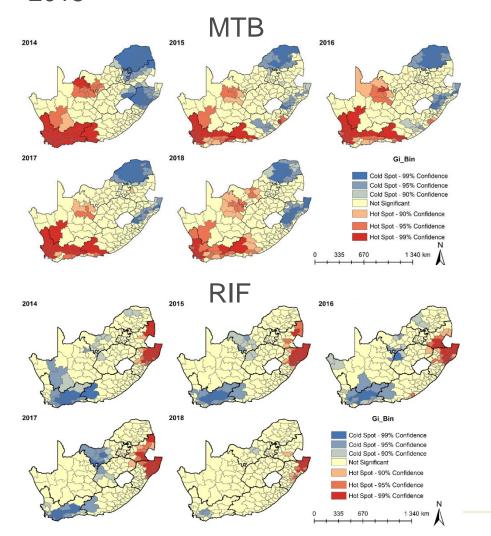
#### **Biometrics Considerations:**

- Database infrastructure (speed of queries)
- Data ownership (security)
- · Identification vs Verification Models
- Mobility

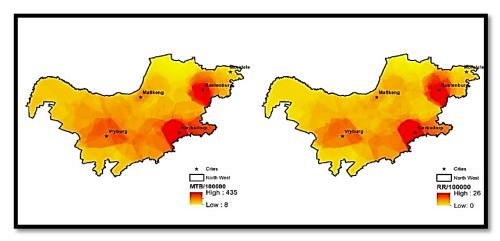
NATIONAL HEALTH LABORATORY SERVICE

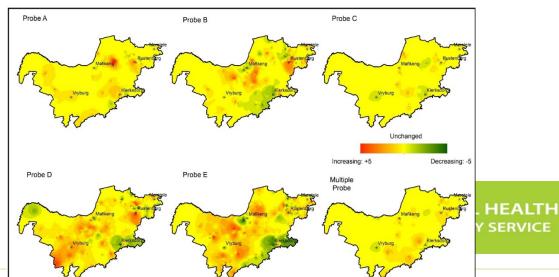
## How is the data used and potential future

Monitoring of the TB Cascade at a National level: Xpert MTB/RIF 2011-2018



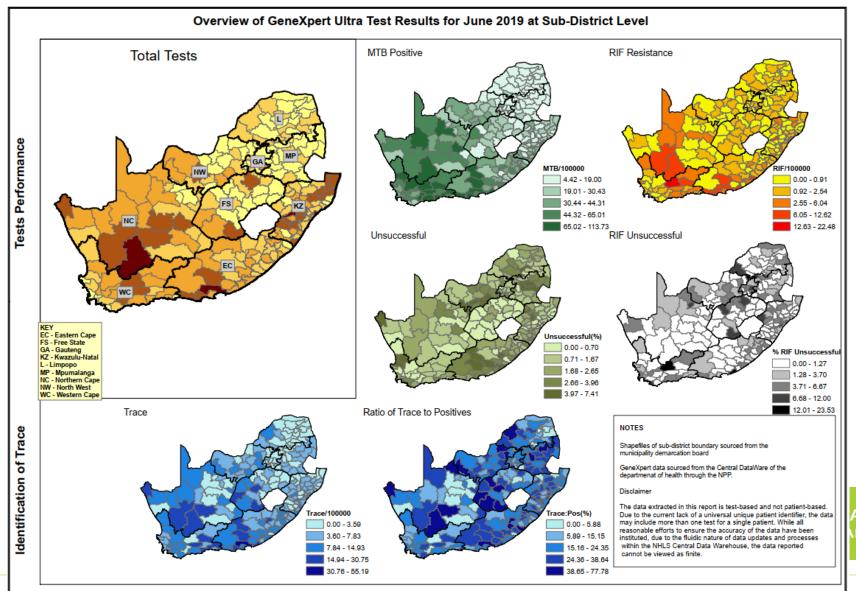
Monitoring of the TB Cascade at a Provincial level: Xpert MTB/RIF 2013-2015 in the North West





## How is the data used and potential future

Operations

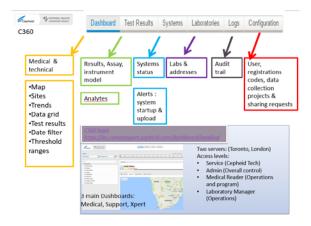


ATIONAL HEALTH
ABORATORY SERVICE

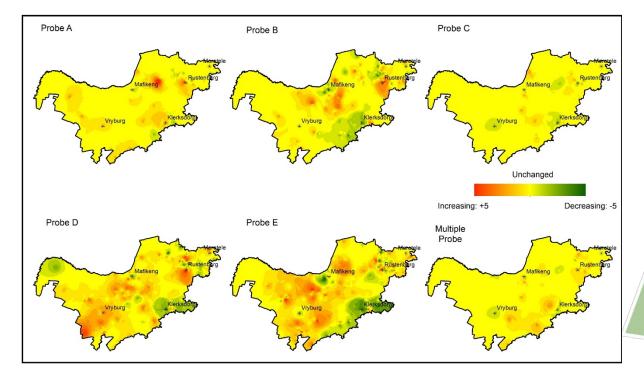
## Continuous quality monitoring



#### Dashboards



#### Molecular granularity



Level 2 Quality monitoring Level 3. Targeted training Level 4: Research analytics

Data curator, storage infrastructure, interfacing, maintenance, hosting, backup

Research concept and design and partnerships

Data access approval and ethics (security) process

Data dictionary, analytics and visualization, tools

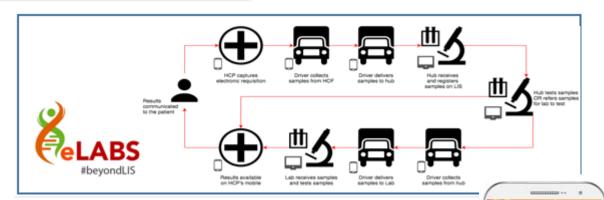
Stakeholder engagement and communications, dashboard access (download speed)

LTH ICE

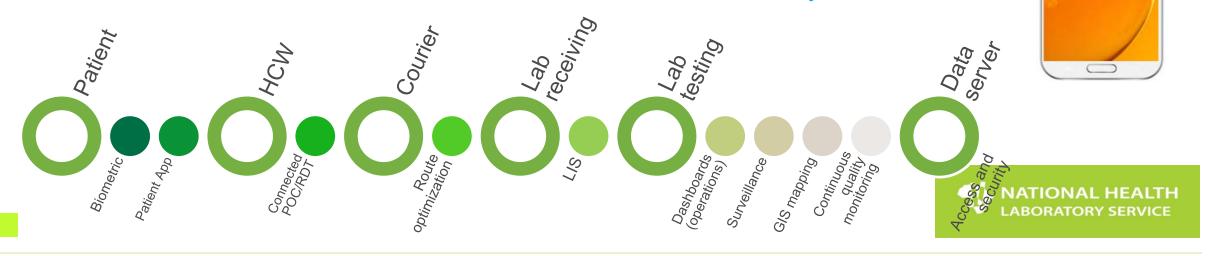
## Expanding the laboratory data backbone

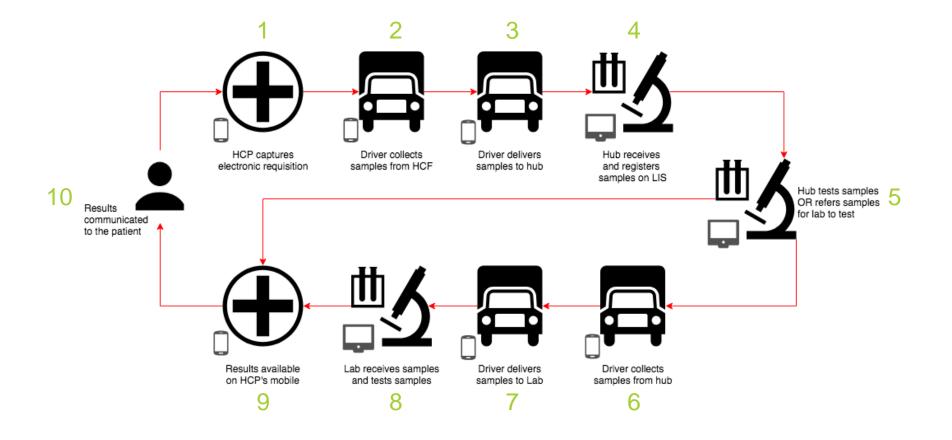


- The "information chain" poor communication between referring clinics and central labs, increases TAT and decreases quality of care.
- Barriers must be addressed
  - Quick and efficient test requests
  - Efficient transport of specimens to central testing labs
  - Standardise communication systems to rapidly transmit results back to clinics (HCW and patients)
  - Workflow automation and monitoring



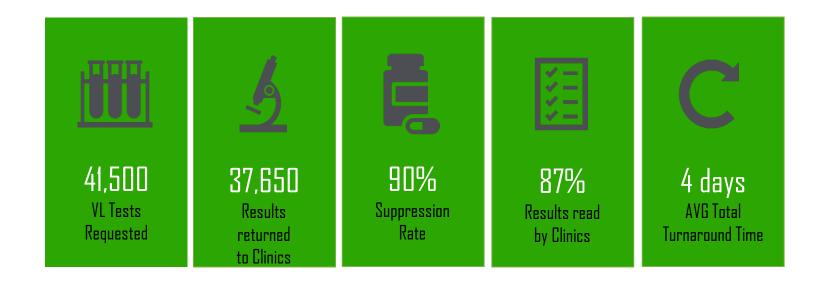
#### eLABS – backbone to a connected laboratory







## Adoption Statistics: South Africa



Project Start Date: October 2018

Reduction in TAT at facility by 62% Registration decrease of 57% Reduction in LSS by 50% Specimen rejection reduced by 1% (2%)% Data Updated: 17 July 2019



## iThen

## NHLS-IT, NHLS-NPP update

September 2019





## "With iThemba, my health is in my hands."

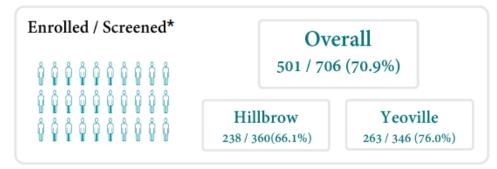
iThemba. It's about time.



## iThemba pilot data (as of June 24th 2019) Enrollments in April, May and June

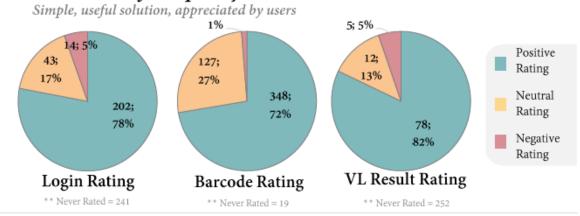
#### 1\*. The Power of Mobile

Harnessing growth of smart mobile access to transform healthcare delivery



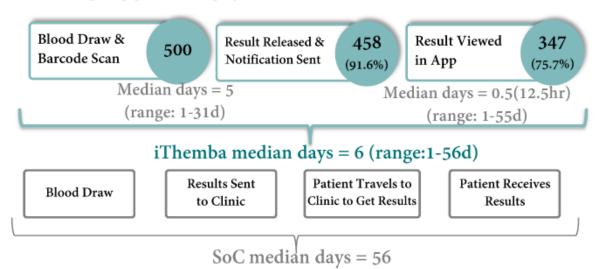
<sup>\*</sup>Some patients are not accounted for in the count due to various IE criteria, only criteria (5-10) are counted in this bucket.

#### 2. The Power of Simplicity



#### 3. The Power of Knowledge

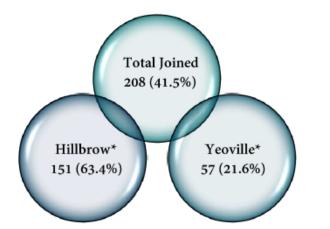
Closing the gap in knowledge of HIV viral load status



(range:10-430d)

#### 4. The Power of Community

Engaging users with virtual communities to provide support



\*Denominator for percent calculation is based on each site's virtual community enrollment











## iThemba

Empowering patients to remain adherent to treatment and engaged in care















## Laboratory as the command centre

- Bi-directional Push and Pull of information
- Continuous real-time monitoring
- Clinical relevancy and footprint expanded by clinical partners
- Priorities to be determined by stakeholders and partner consultations
- Complexity will increase beyond the proof of concepts
- Platform for innovation and social entrepreneurship
- Significant investment required for maintenance



## Considerations

- The CDW storing all public sector laboratory results is indeed a national treasure and needs to be resourced
- The value of the laboratory for individual patient management has always been relatively clear.
- New roles:
  - The value of aggregated laboratory data has been demonstrated in the projects presented today: programmatic value
  - The role of the laboratory in proactive enhancement of linkage to care via solutions such as mHealth, webview etc is in its infancy requires investment
  - Facilitates a platform to support POCT initiatives
  - Monitoring of analyzer performance from a distance



## Acknowledgements

- National Department of Health (HIV and TB cluster)
- NHLS National Priority Program
- Department of Molecular Medicine and Haematology, Wits University R&D team
- NHLS, National Priority Program, TB/HIV Expert working groups
- Funders (specifically CDC, Bill and Melinda Gates foundation, NIH/ACTG, USAID, MRC/Newton)
- Clinical partners Right to care, Aurum, TB/HIV care, WRHI, PHRU
- Research collaborators: HERO (BU),
- Commercial collaborators (Cepheid, Abbott, Roche, Hain, BD)
- Innovators







































