



Western Cape  
Government  
FOR YOU



UNIVERSITY OF CAPE TOWN  
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

# Population-based epidemiology in the era of data science and routine health data

## Western Cape Province of South Africa – Health example

Andrew Boulle

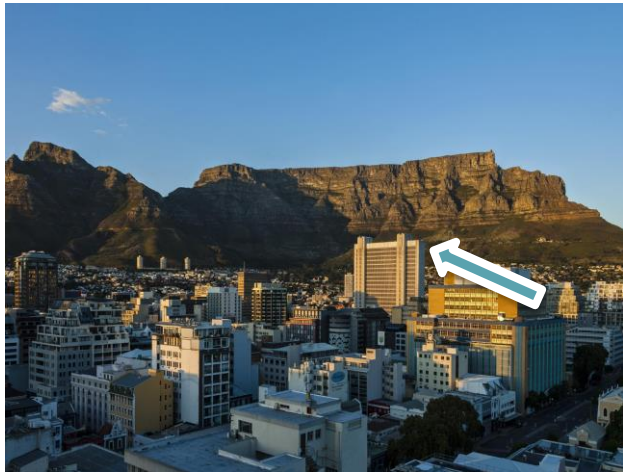
Western Cape Department of Health and Wellness  
University of Cape Town

# Approach to this session

---

1. Background to the Western Cape health service and integrated health data
1. Data governance and how epidemiological research is supported
1. Reflections and challenges

# Western Cape Province: Population context



**7.5 million**

Population (75%+ using public sector)

**550,000**

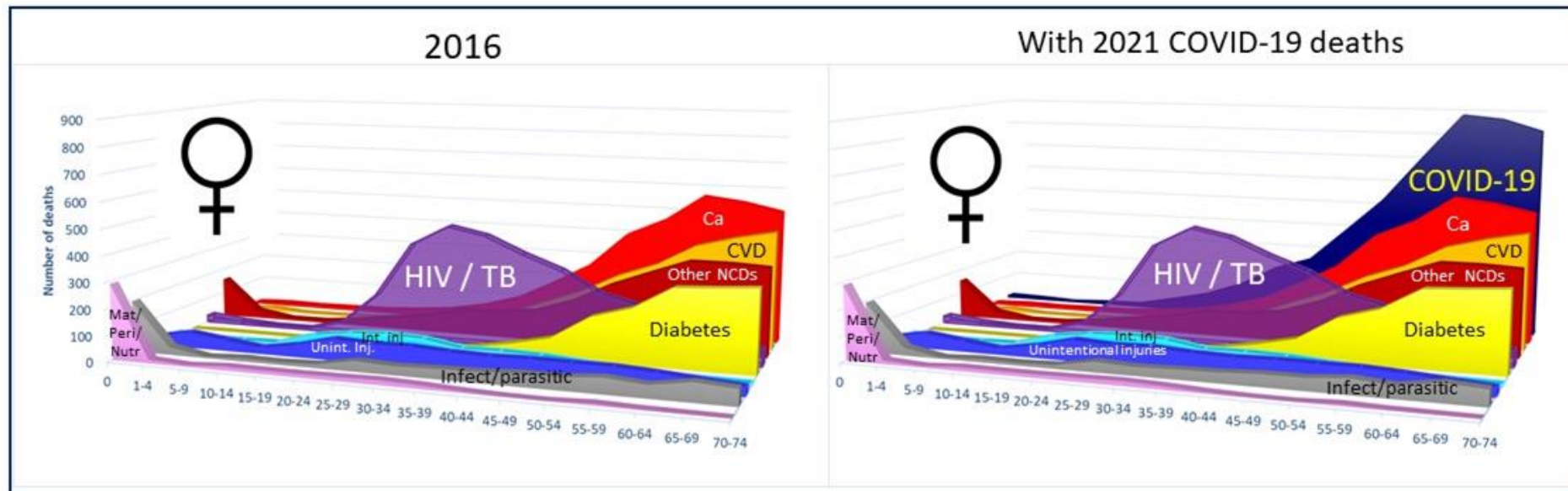
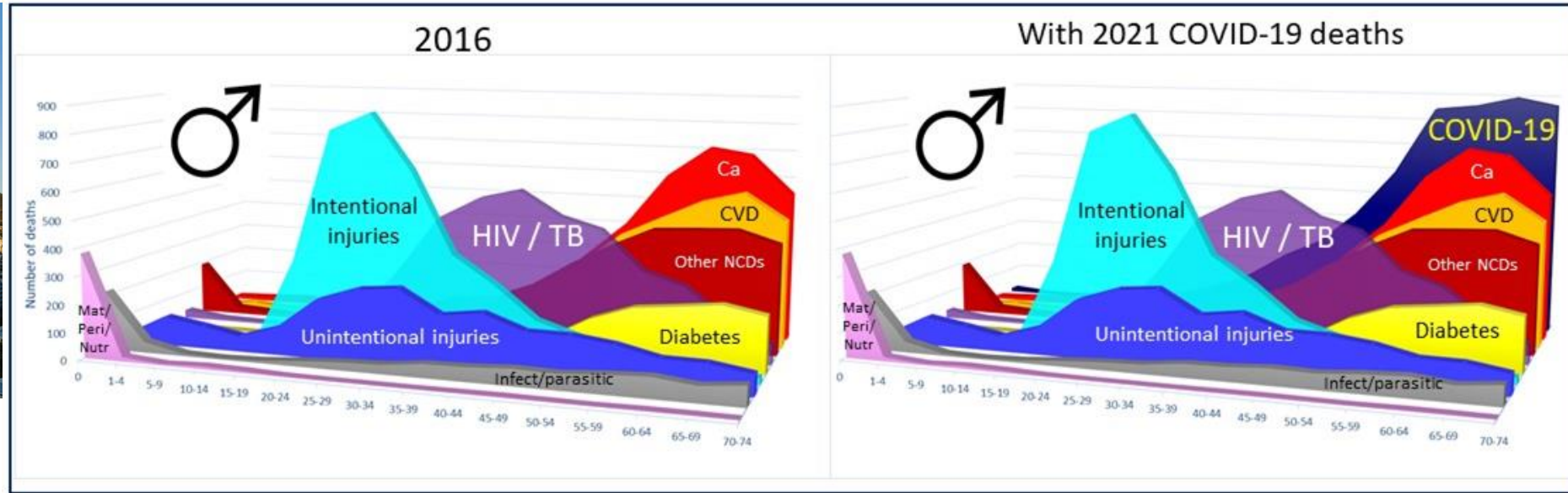
people living with HIV

**50,000**

people diagnosed with TB each year

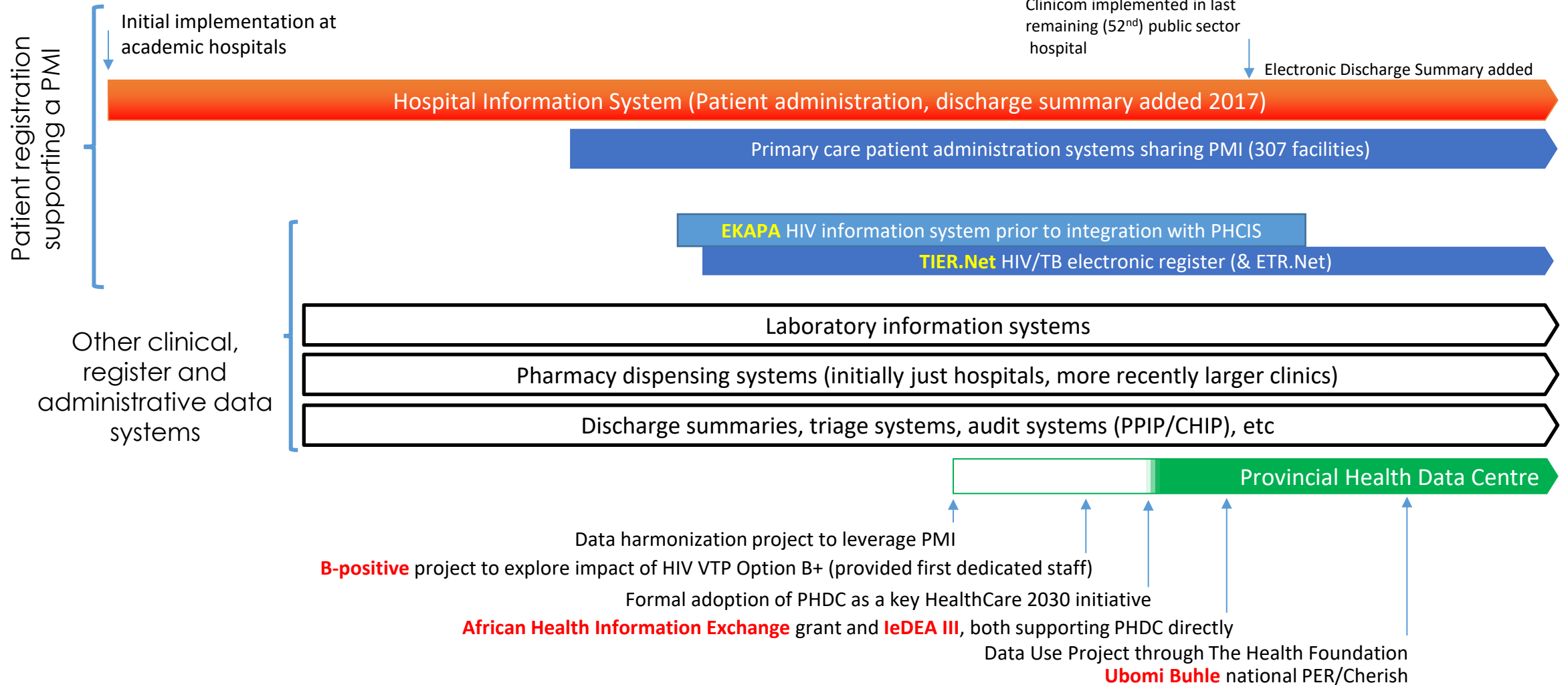
**30,000**

Excess deaths during COVID-19  
(430/100k)

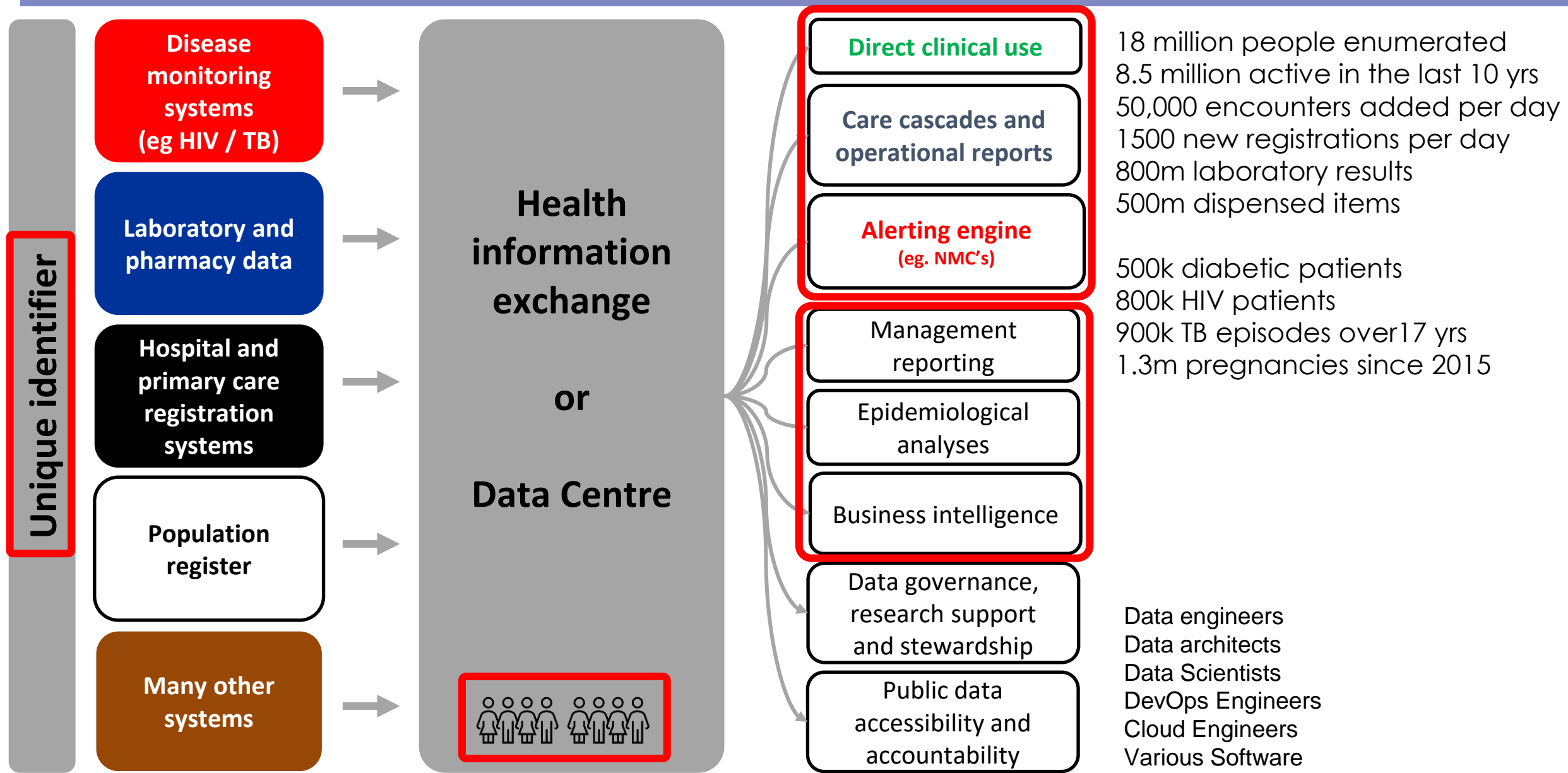


# Timeline for key digital health systems supporting data integration

1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022



# A Provincial Health Data Centre



# Context specific health information exchange considerations

## (1) An imperfect unique health identifier with imperfect deduplication



Western Cape Government | Single Patient Viewer | Demo

Female, \*\* X | Search Folder

**CAUTION:** As the user, you are responsible for verifying all presented SPV data before using them in any clinical decisions. Absence of data may not mean that the user did not access services, but could be due to their visits or rec...

Folder #	ID No	First Name	Last ...	Address	Date of Birth	Dat...	Age	Gen...	Link Patients	Group Episodes
[Redacted]	[Redacted]	[Redacted]	[Redacted]	[Redacted]	[Redacted]	[Redacted]	**	F		
[Redacted]	[Redacted]	[Redacted]	[Redacted]	[Redacted]	[Redacted]	[Redacted]	**	F	ON OFF	ON OFF
[Redacted]	[Redacted]	[Redacted]	[Redacted]	[Redacted]	[Redacted]	[Redacted]	**	F	ON OFF	ON OFF

# Context specific health information exchange considerations

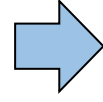
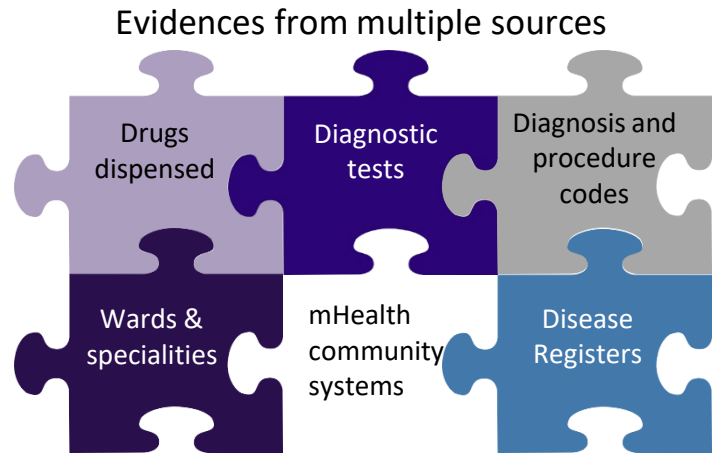
## (2) Incomplete and inconsistent clinical coding

---

- Not an insurance system so **no claims-based coding**
- The introduction of an electronic discharge summary completed by clinicians has improved hospital coding but is still **not universally used and often incomplete or inaccurate**
- **No coding of primary care visits** (though e-scripting will improve this in near future)



# Phenotype algorithms for key health conditions

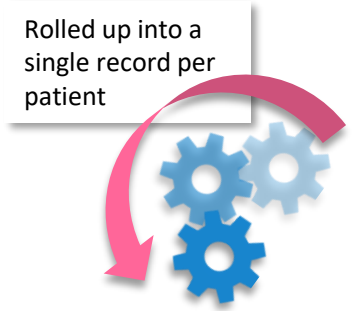


Categorised based on specificity to data point of interest

Outcome	Evidence of an outcome for the episode
High confidence	Strong enough evidence to start an episode
Weak-Moderate evidence	Able to start an episode, however more evidence is required to improve
Supporting only	Can only be appended to existing episode but not start one

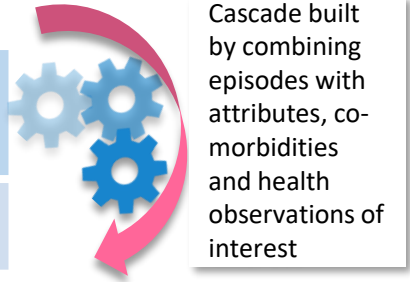
## Evidences

Evidence	Source	Category	Date
Has Random glucose greater than 11.1 mmol/l	Labs	High confidence	2015-01-23
Dispensed drug for treatment of diabetes	Drugs	High confidence	2015-04-03
Has CHW report of diabetes	Community	Weak-Moderate	2016-08-16
Has ICD10 code indicating diabetes	Diagnosis codes	Weak-Moderate	2015-06-16



## Diabetes episode

Patient	First evidence Date	Last Contact	Evidence list	Treatment start date	First evidence facility	Treatment facility	Last evidence facility	Confidence
xxx	2015-01-23	2017-06-16	Random Glucose, drugs, community, ICD10	2015-04-03	Hosp A	Clinic A	Clinic B	0.95 High confidence



## Diabetes cascade

First evidence date	First evidence Date	Last Contact	Treatment start date	Last drug date	First evidence facility	Treatment facility	Last evidence facility	Last oral drug	Last insulin	Last HBA1c date	Last HBA1c value	Last eGFR date	Last eGFR value	Co-morbidities	Evidence list	Confidence
2015-01-23	2015-01-23	2017-06-16	2015-04-03	2017-06-16	Hosp A	Clinic A	Clinic B	2017-06-16	2017-06-16	2016-08-11	9.3	2016-08-11	55	Hypertension (2015)	Random Glucose, drugs, community, ICD10	High



---

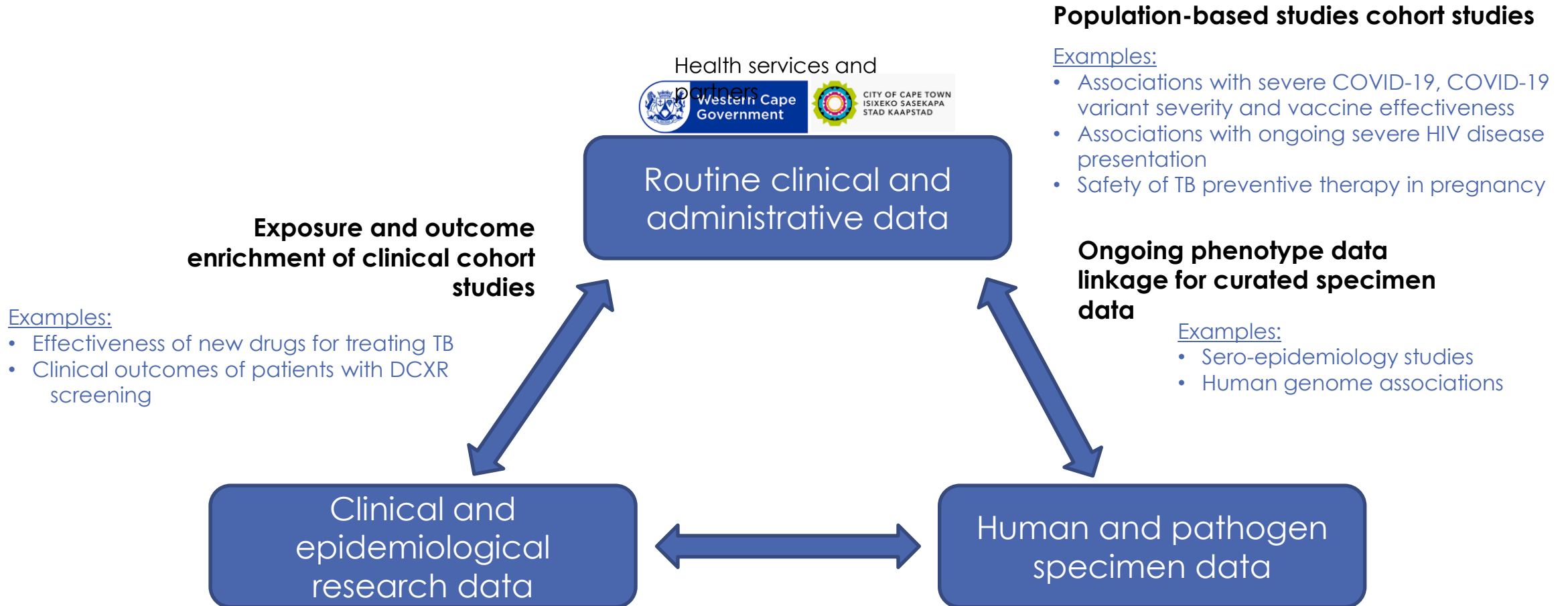
**Data Governance  
and support for epidemiological research**

# Data governance and research data management

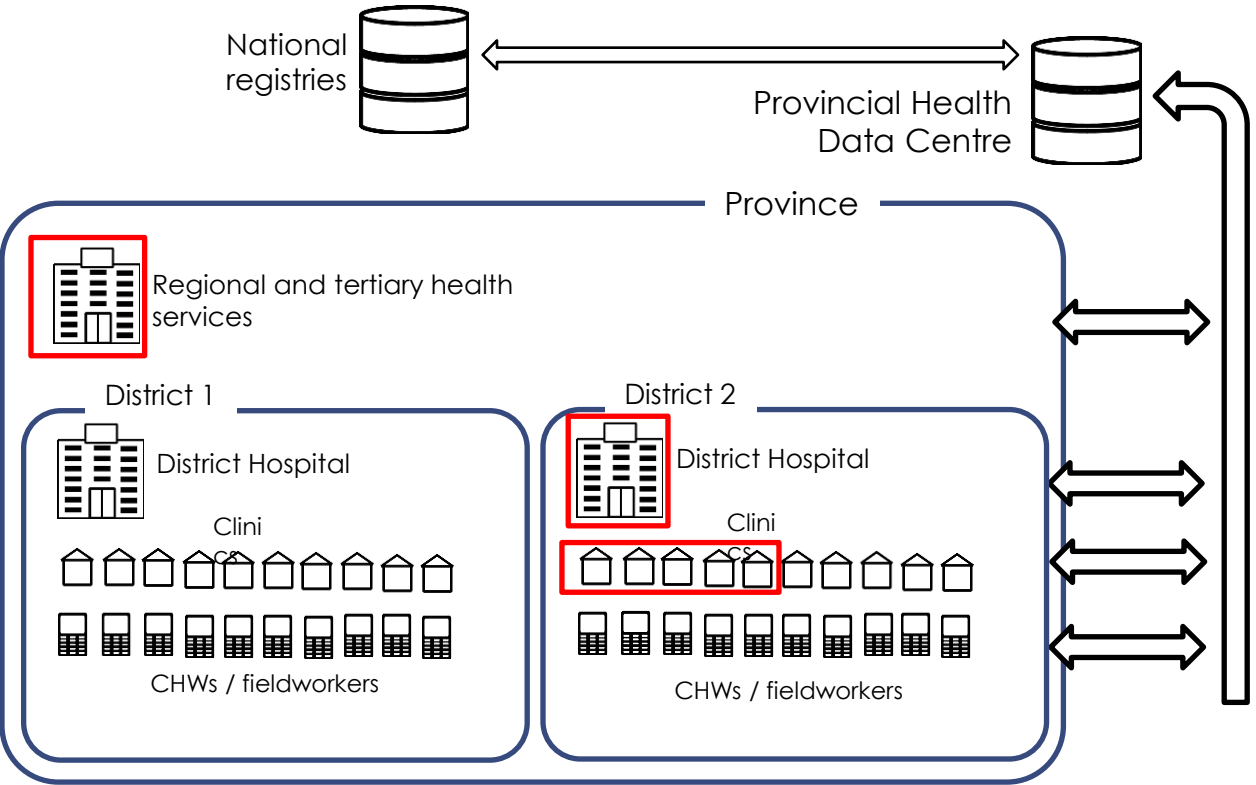
---

- **Single provider public health system** with implicit social contract to use identifiable data only for the purposes for which it was shared
- The Western Cape Provincial Health Data Centre is **not a research organisation**
  - It is essential that PHDC is not confused with being a research entity
  - But, the PHDC **can support research** (internally and external to the health department) through facilitating research data access
- Can only share data for research (i.t.o. health and general privacy legislation)
  - If explicit consent is provided, or
  - If de-identified, and progressively being implemented, if all the conditions of the 5-safes are in place
- Over 70 articles reference the PHDC as a data source for their study
- **No routine access to non-health data** for exploration of social determinants on health, or evaluation of interventions or outcomes outside the health services
  - Currently there is no entity in the province or nationally which is explicitly sanctioned to securely hold all identified data across service and administrative domains (CIDACS model)

# Supporting epidemiological research



# Enriching the service data held by the WC Department of Health and Wellness (1)



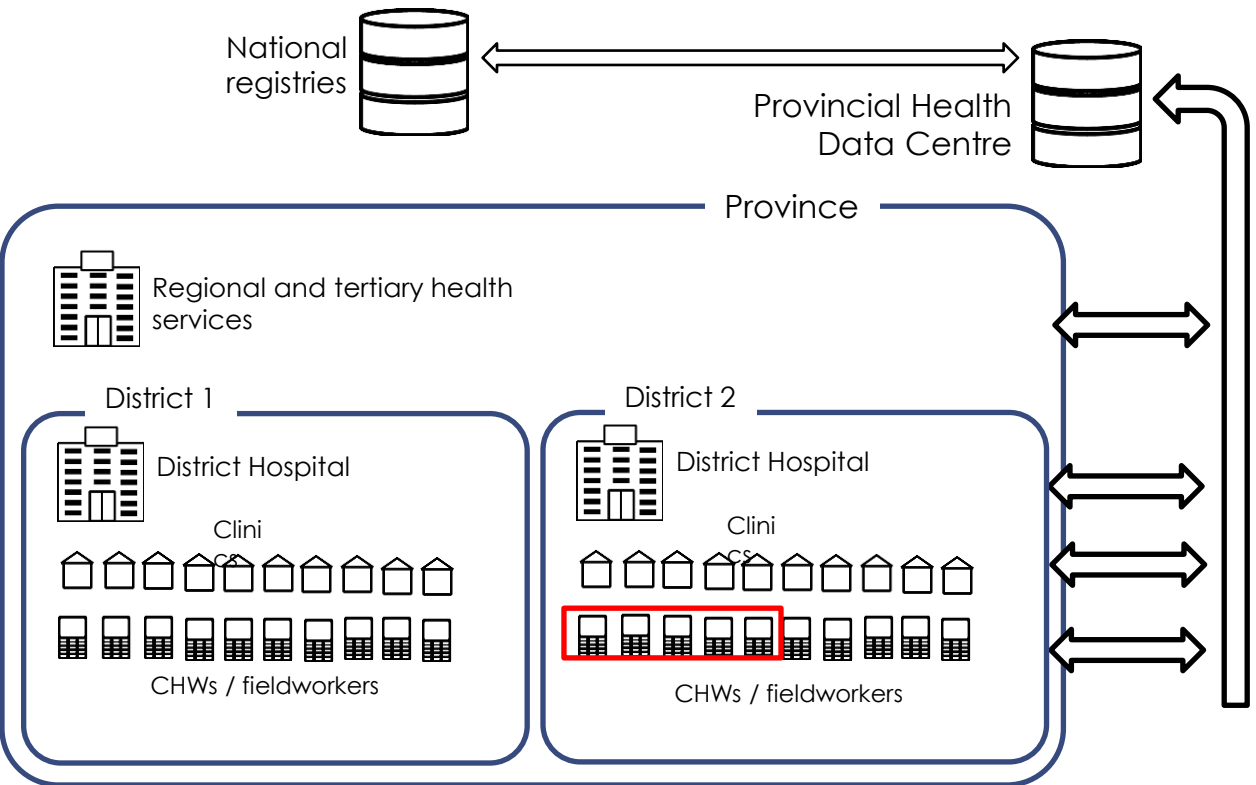
**Digitising additional routine service data**  
(Not consented)

Open access Cohort profile

**BMJ Open Cohort profile: the Western Cape  
Pregnancy Exposure Registry (WCPER)**

- Standard platforms
- Data are available in the shared health record to support clinical care

# Enriching the service data held by the WC Department of Health and Wellness (2)



## Collecting additional non-service data (Consented)

Cape Town **Health and Demographic Surveillance Site** has been designed to be implemented through the service delivery platform by community health workers.

- Standard platforms
- Facilitates linkage of community and service data by design



**SAPRIN**  
POPULATION-BASED SCIENCE



CAPE TOWN SYSTEMATIC  
HEALTHCARE ACTION  
RESEARCH PROJECT



# Reflections and challenges

# Data governance and research data management

---

## Core data quality and coverage issues

- Unique identification challenges
- Providing provenance on how records were linked can enable analysts to explore impact of trusting and not trusting linkages
- Mortality data – in spite of a good vital registration system, the cause of death data are not available to the health department
- Ascertainment of health conditions and events which do not have strong laboratory or medicine signatures

## Inherent limitations of a service-sourced cohort

- People could have left through unascertained mortality and migration, or accessed services outside the network (e.g. private sector)
- Absence of diagnosis does not mean absence of event
- HDSS and surveys like DHS can help explore these issues

## Linking other exposures, risk factors and outcomes (e.g. environmental, socio-economic)

- Assigning values based on where people live can be considered, but poor-quality address data can limit the granularity
- Linking to non-health data sources through trusted third parties, privacy-preserving linkage, in absence of legislated mandates

Resources – limited funding for dedicated staff to support research



# Questions?



## Acknowledgements

Department of Health and Wellness, WCG, for the many colleagues who have supported the source systems which feed into the PHDC, Health Intelligence colleagues, clinical colleagues, and patients

---

**Examples of support for research  
(Extra slides at the end, only if time allows)**

# 1. Population-based cohorts

– TB and pregnancy combined

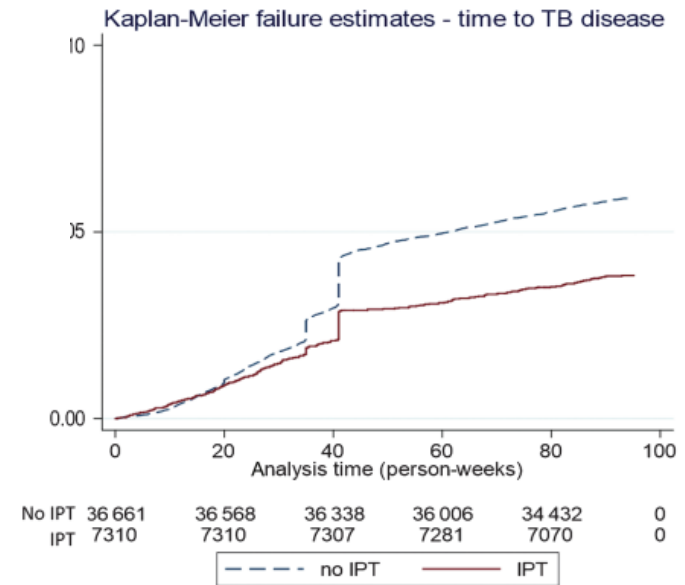
*Clinical Infectious Diseases*

MAJOR ARTICLE



## Safety and Effectiveness of Isoniazid Preventive Therapy in Pregnant Women Living with Human Immunodeficiency Virus on Antiretroviral Therapy: An Observational Study Using Linked Population Data

OR for adverse pregnancy outcomes of 0.83



# 1. Population-based cohorts - HIV

## Outcomes after two years of providing antiretroviral treatment in Khayelitsha, South Africa

*AIDS* 2004, **18**:887–895

287 patients

*Clinical Infectious Diseases*

SUPPLEMENT ARTICLE



## The Continuing Burden of Advanced HIV Disease Over 10 Years of Increasing Antiretroviral Therapy Coverage in South Africa

Anderson K et al. *Journal of the International AIDS Society* 2024, **27**:e26235  
<http://onlinelibrary.wiley.com/doi/10.1002/jia2.26235/full> | <https://doi.org/10.1002/jia2.26235>



RESEARCH ARTICLE

## Factors associated with vertical transmission of HIV in the Western Cape, South Africa: a retrospective cohort analysis

Now we have a province-wide HIV cohort of over 800,000 people who have ever been enumerated with HIV in the Province. In year 18 of NIH funding through the leDEA Southern Africa cohort collaboration

# 1. Population-based cohorts – COVID-19

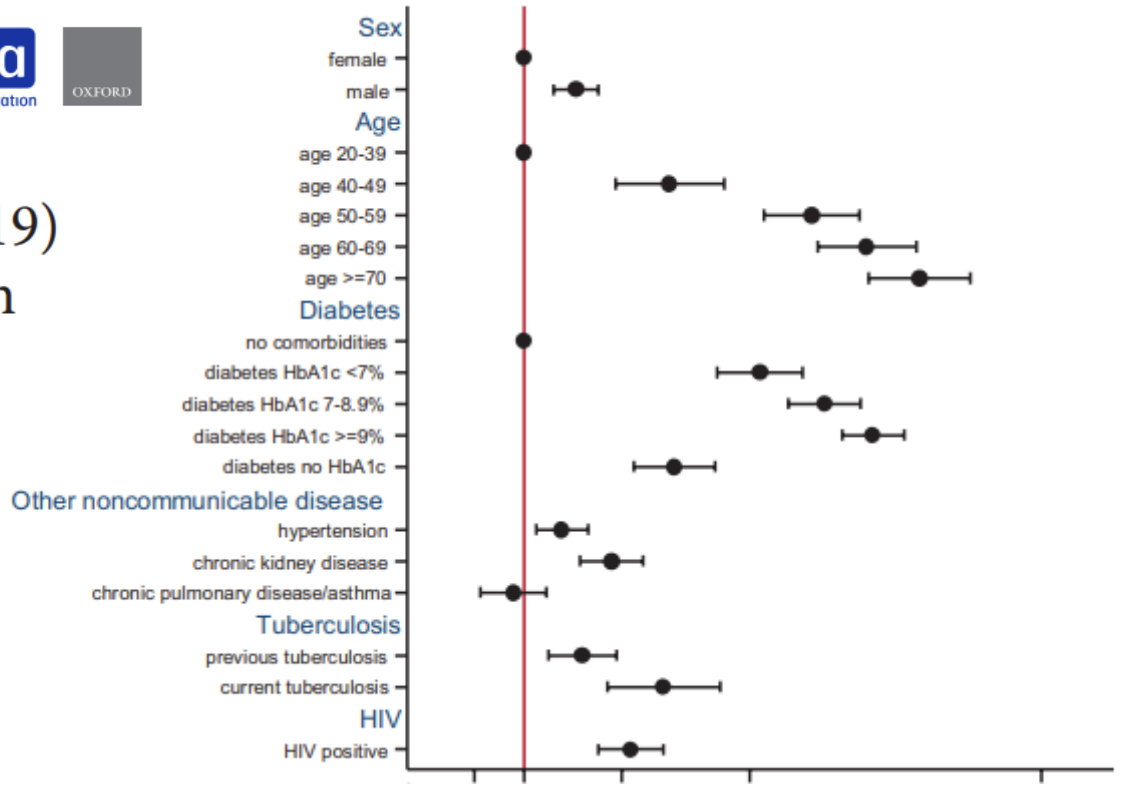
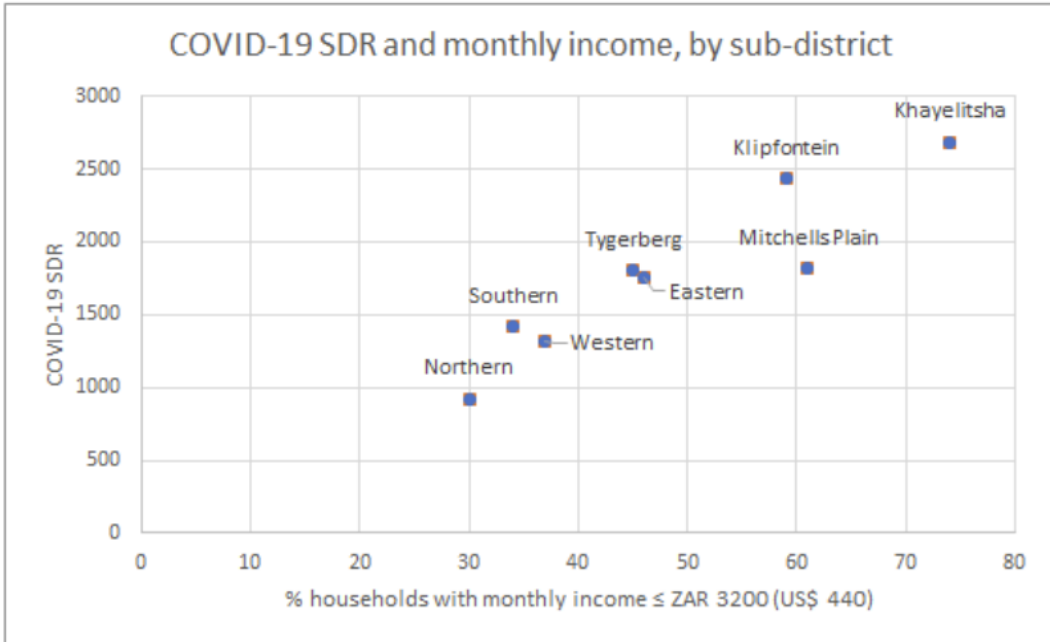
Clinical Infectious Diseases

MAJOR ARTICLE



## Risk Factors for Coronavirus Disease 2019 (COVID-19) Death in a Population Cohort Study from the Western Cape Province, South Africa

Western Cape Department of Health in collaboration with the National Institute for Communicable Diseases, South Africa



Gates Open Research

Gates Open Research 2021, 5:90 Last updated: 23 MAR 2022



RESEARCH ARTICLE

**Higher COVID-19 mortality in low-income communities in the City of Cape Town – a descriptive ecological study [version 1; peer review: 2 approved]**

## 2. Supporting investigator-led primary research

RESEARCH ARTICLE

LMIC-PRIEST: Derivation and validation of a clinical severity score for acutely ill adults with suspected COVID-19 in a middle-income setting

PLOS ONE | <https://doi.org/10.1371/journal.pone.0287091> June 14, 2023

Frigati LJ et al. *Journal of the International AIDS Society* 2021, **24**:e25671  
<http://onlinelibrary.wiley.com/doi/10.1002/jia2.25671/full> | <https://doi.org/10.1002/jia2.25671>



RESEARCH ARTICLE

**Tuberculosis infection and disease in South African adolescents with perinatally acquired HIV on antiretroviral therapy: a cohort study**