

# Probabilistic record linkage approaches to enhance HIV treatment data for research: Is it feasible?

Dickman Gareta<sup>1</sup>, Dorina Onoya<sup>2</sup>, Kobus Herbst<sup>1,3</sup>, Jacob Bor<sup>4</sup>

<sup>1</sup> Africa Health Research Institute, South Africa <sup>2</sup> University of the Witwatersrand, South Africa <sup>3</sup> South African Population Research Infrastructure Network, South Africa <sup>4</sup> Boston University, USA

## 1. Background

- South Africa runs the largest HIV program in the world – 7.5 million PLWH and 5.1 million on ART in 2020
- Massive ART scale up in the past decade
- Critical to assess the impact of the ART scale up on HIV treatment outcomes
- HIV data collected and stored in disparate data systems
- Linking HIV data remains a challenge

## 2. Objective

- To build an enhanced HIV care and treatment database for research and M&E purposes
  - a. To develop probabilistic *linkage and deduplication* of HIV treatment and lab data
  - b. Validate performance using SA national IDs with a novel approach for bias-correction

## 3. Methods

### Data Sources

#### HIV data

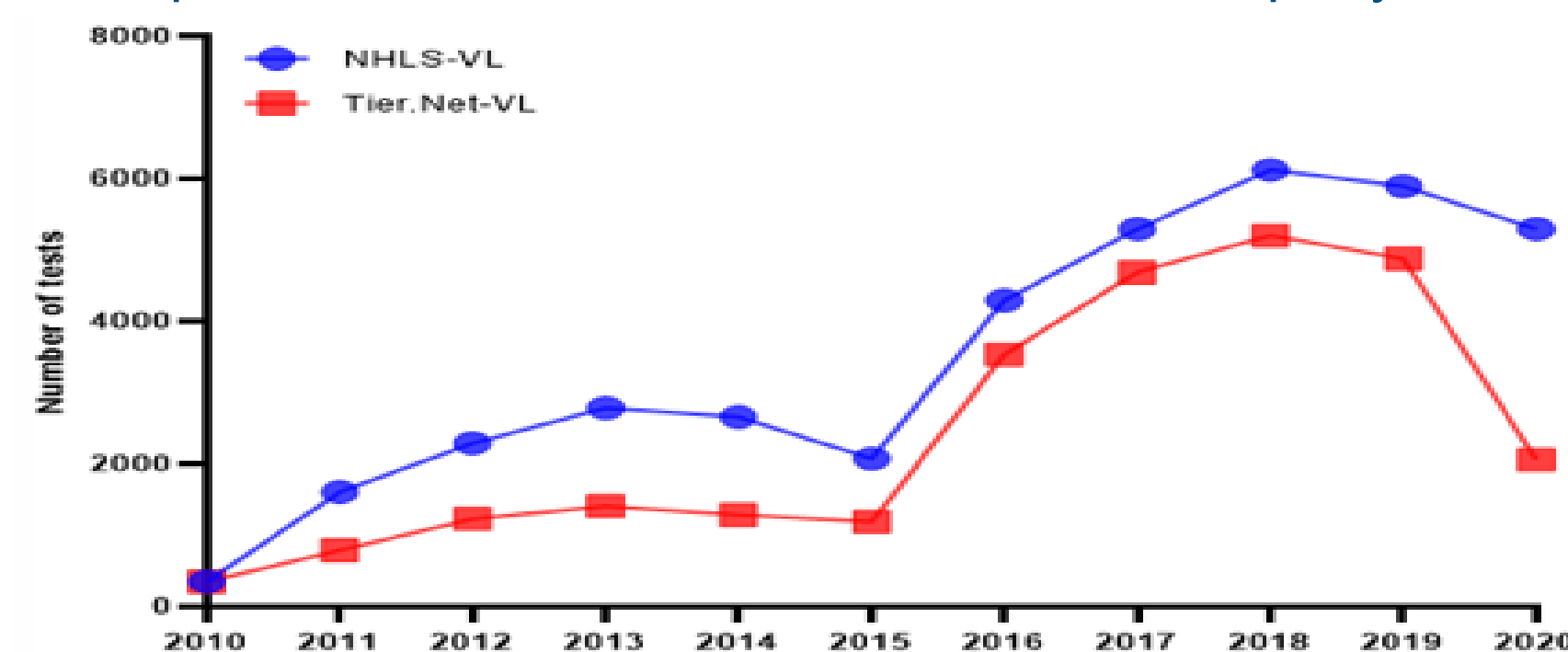
- **TIER.Net** is used to capture HIV/ART care and treatment in SA
- Electronic patient management system for M&E
- Records are captured in a non-networked computer
- Implemented in uMkhanyakude district in 2013

#### Laboratory data

- All public health facilities send samples to **NHLS**
- NHLS is a sole provider of diagnostic pathology lab
- Implemented in KwaZulu-Natal in 2010
- AHRI has MOA with Department of Health and NHLS to access ART and laboratory data for 17 clinics in Hlabisa sub-district

## 4. Motivation

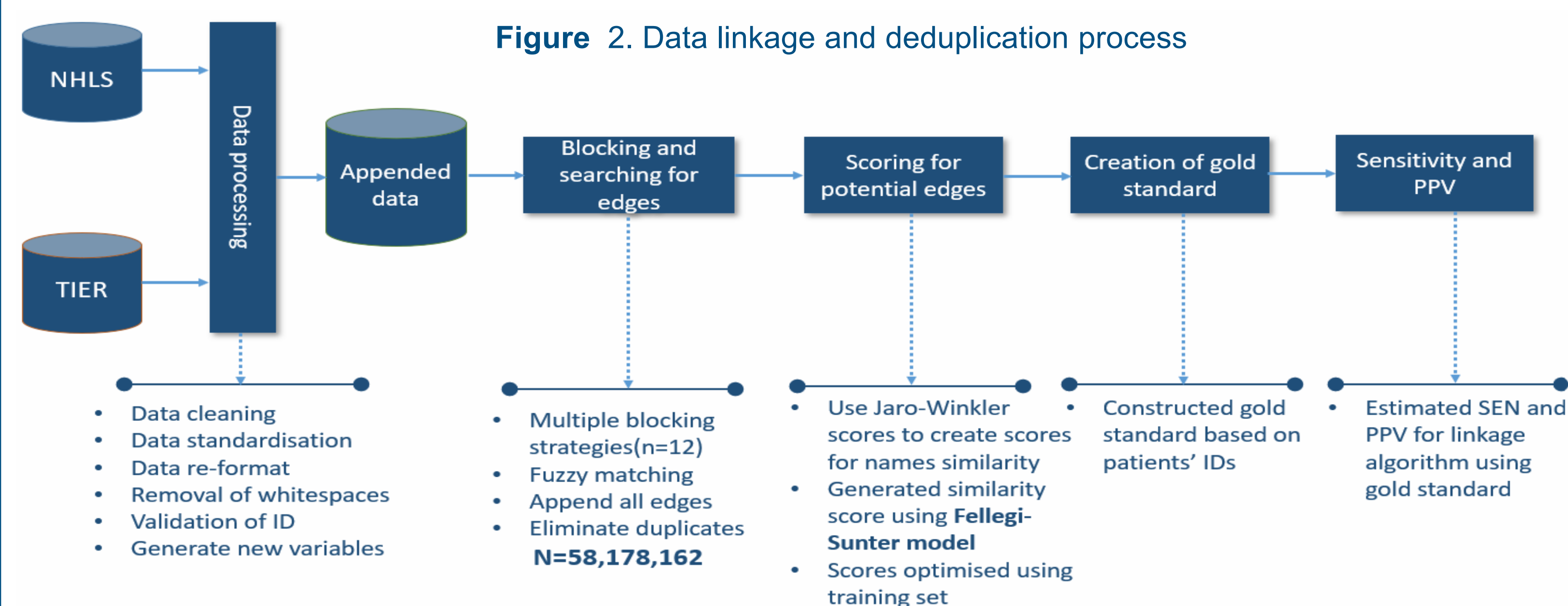
Figure 1. Comparison of NHLS and TIER VL tests done per year at clinic A



Challenge: This can lead to underestimation of viral load tests

## 5. Record linkage process

Figure 2. Data linkage and deduplication process



## 6. Results

Figure 3. Total similarity score distribution

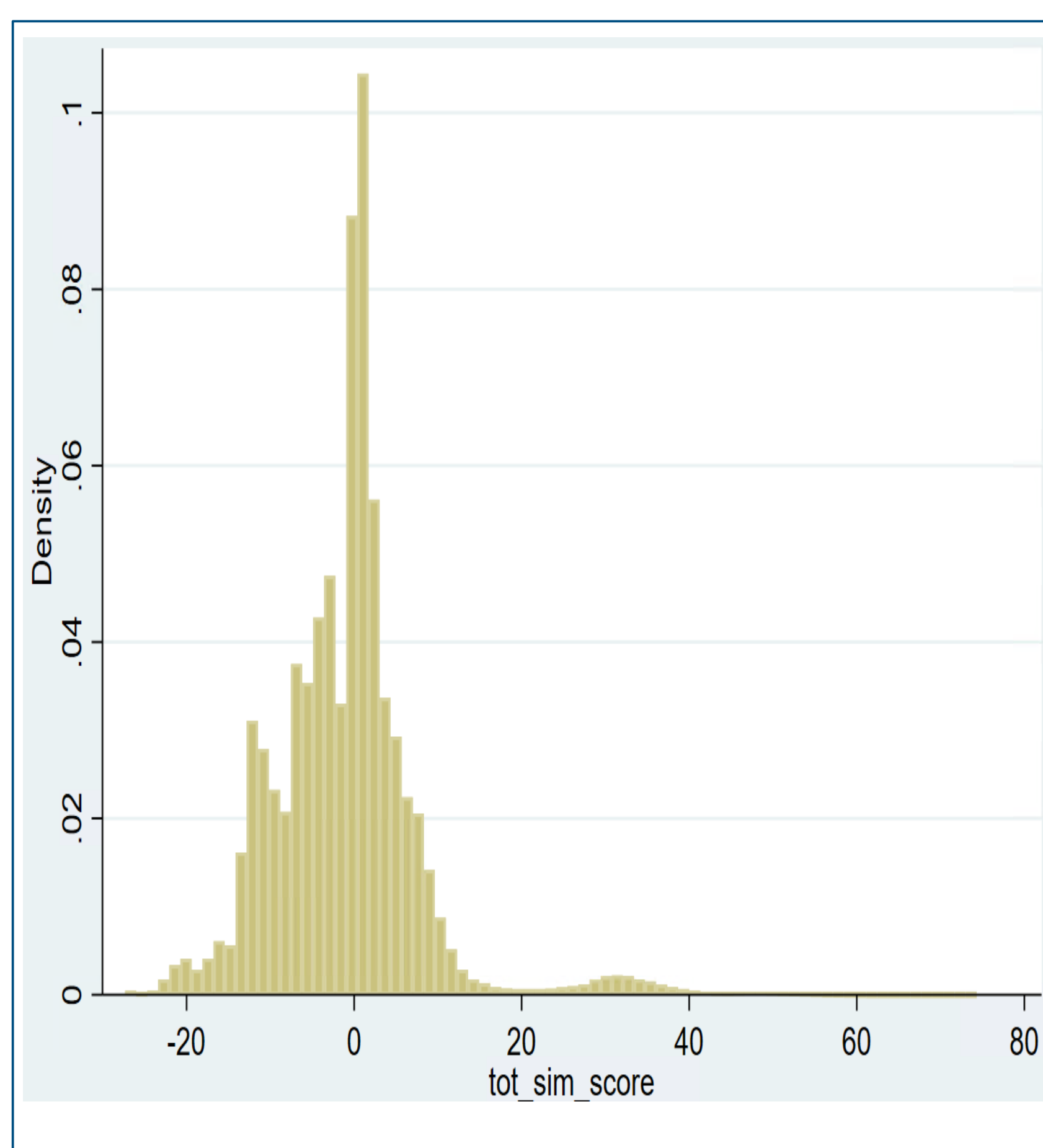


Figure 4. Probability that national IDs match by total similarity score

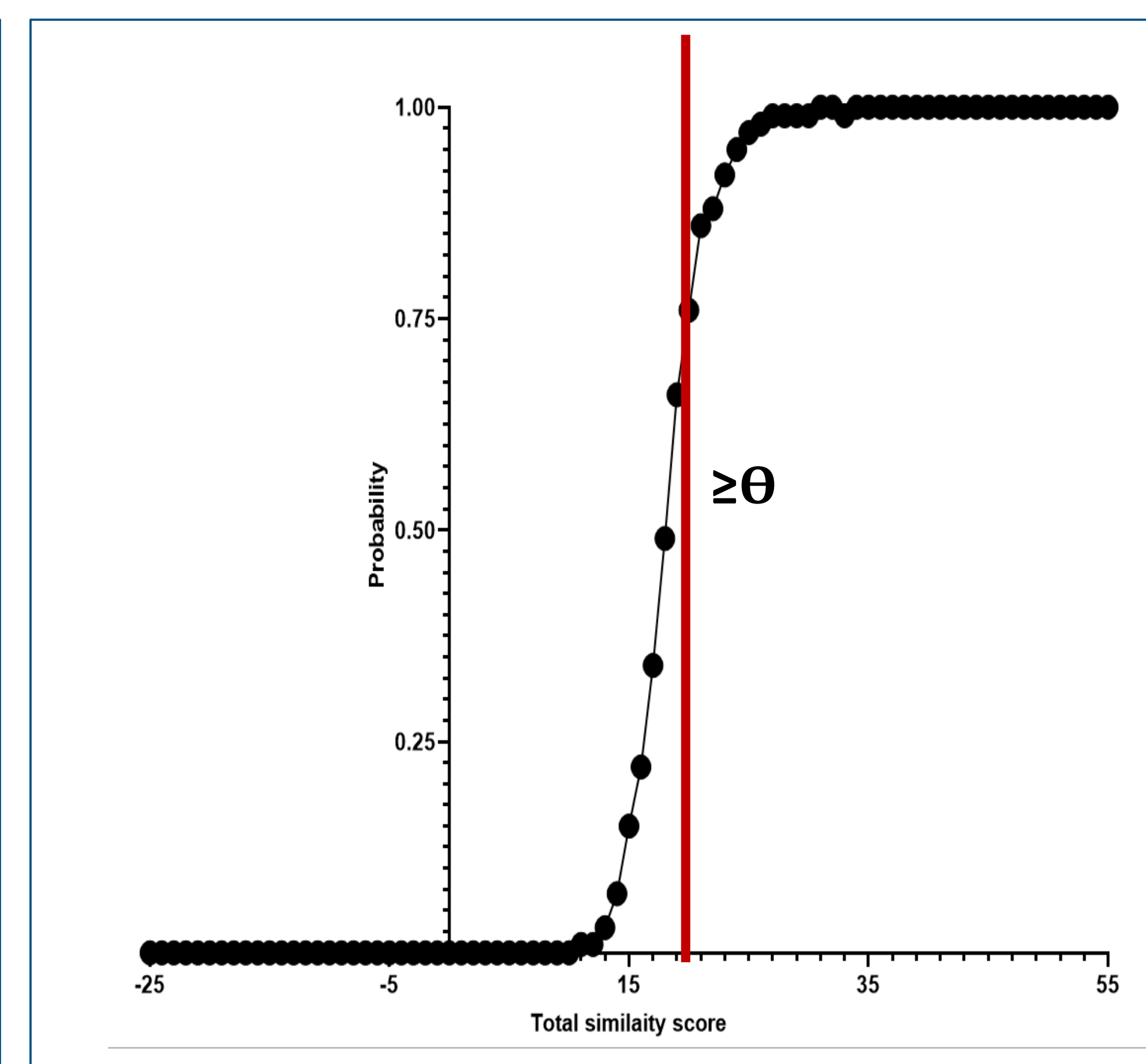
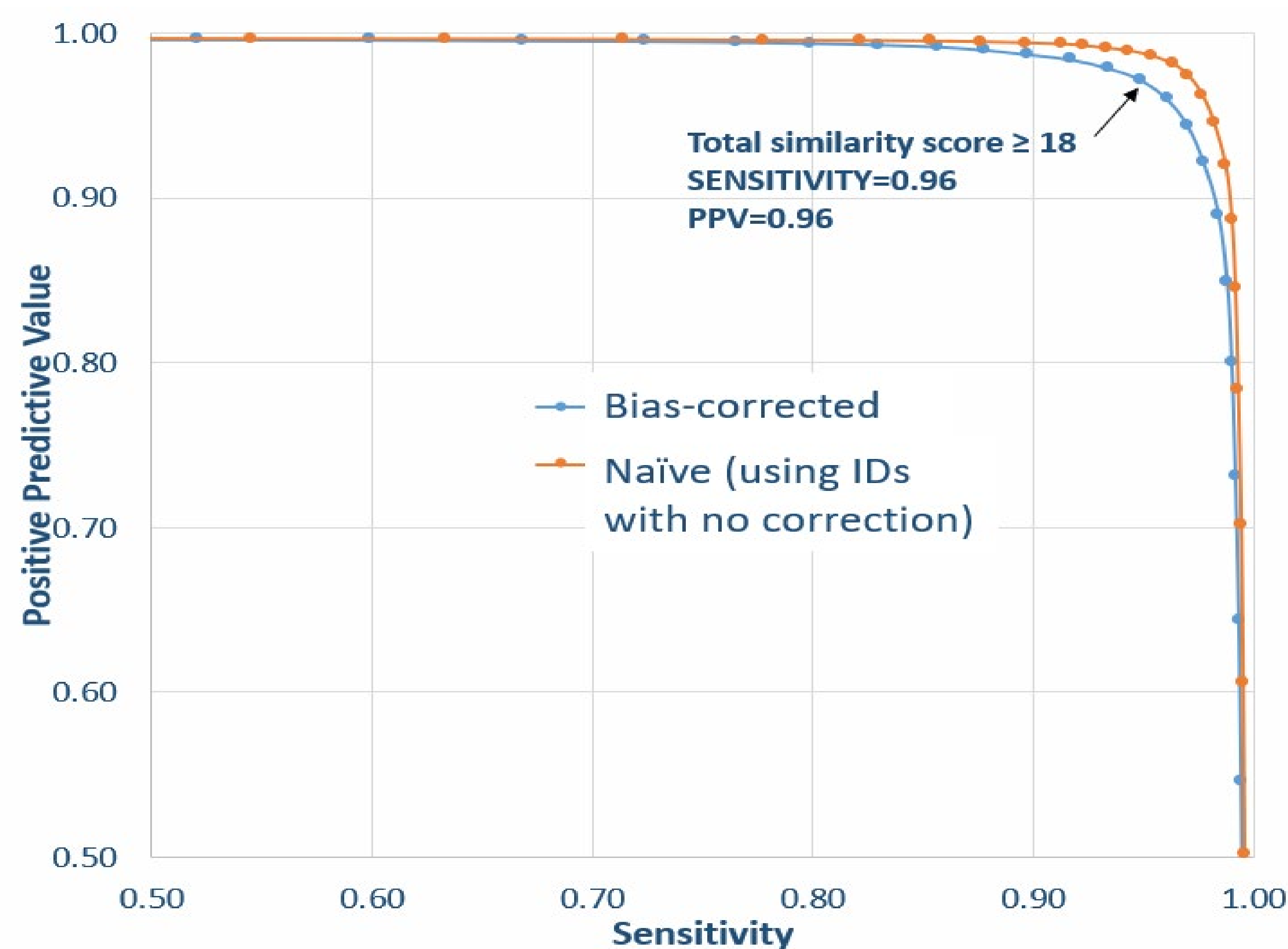


Figure 5. Sensitivity and positive predictive values



Our approach provides a scalable template for sub-district level record linkage of HIV and lab data in South Africa

## 7. Conclusion

- Its possible to replicate our approach to any sub-district in South Africa
- Possible to link ART and lab databases with high accuracy in settings with over 300,00 records contributed by over 50,000 people
- Our validation exercise did not require a manual review
- Our approach addressed bias due to informative missingness
- Offers excellent opportunity to fully understand the impact of UTT on HIV treatment outcomes