



**AFRICA CENTRE**  
For Population Health



# Linking HDSS data to health services data: Challenges and opportunities

Dickman GARETA, Siyabonga NXUMALO and Kobus  
HERBST

**Africa Centre for Population Health**

INDEPTH Scientific Conference, 11-13 Nov 2015



**INDEPTH Network**

**welcome**trust

# Overview

- **Background**
- **Research setting**
- **Data sources**
- **Methods**
- **Data linkage process**
- **Results**
- **Challenges**
- **Opportunities**
- **Conclusion**

## Background

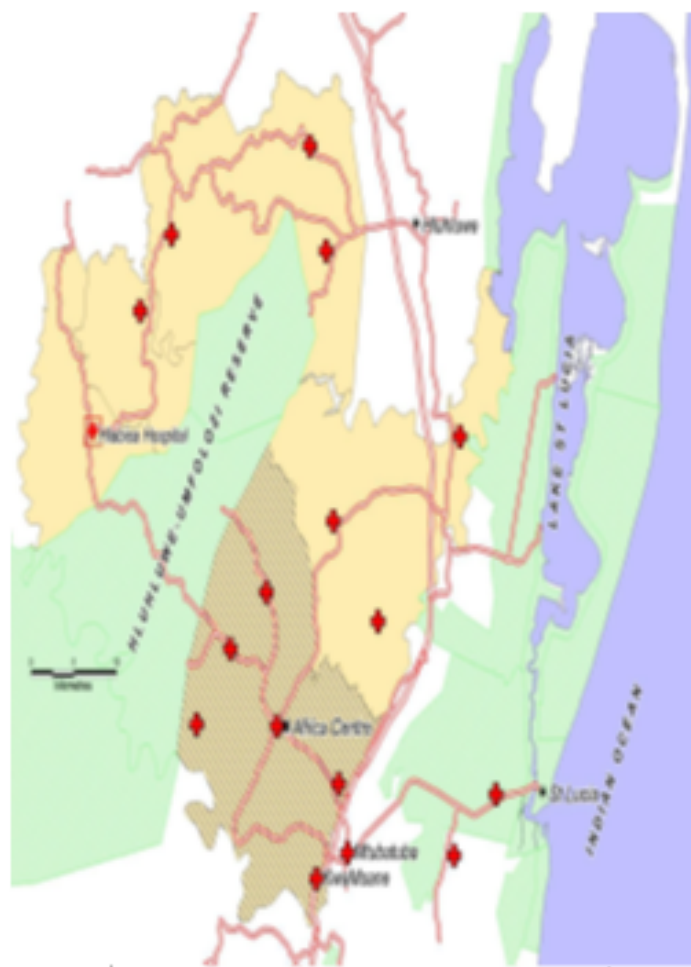
- HDSS collect huge volume of data
  - adequately enough?
- Increasing interest in using data linked from multiple sources
  - 6-fold increase in the last two decades (Bohensky *et al*, 2010)
- Data linkage centres have been established globally
  - >60 in Australia, 23 in Canada, 5 in UK, 1 Germany, 1 New Zealand, 2 Austria and 7 in Brazil

## Background

- Data linkage (DL) provides:
  - Wide coverage and scope of data
  - More complete info of individuals
  - Most cost-effective ways of supporting research (WT Report,2015)

## Setting

- High poverty
  - Unemployment (~67% ,2010)
- High disease burden
  - 28% adult HIV prevalence
- 200-bed district hospital outside the DSA
- HIV care is provided in the sub district through 17 primary health care clinics



## Data sources

- HDSS research data
- Hospital information system
- Electronic ART record system(ARTemis/Tier.Net)
- Electronic routine lab results(NHLS)
- Electronic pharmaceutical data (iDART)
- Electronic TB Register (ETR) (in progress)

There is a need to link all these disparate sources - that is complex and time consuming and you need an effective and efficient work flow to do it routinely

## Data linkage levels

- ART data (Tier.net) with ART data (Tier.net)
- ACDIS and ART data
- ACDIS and hospital data

**Data from the same individual can appear in multiple systems and the ability to integrate this data depends on accurate record linkage**

# Deterministic data linkage

Largely use national id to link data sets

-77% (Resident & age>15) have valid

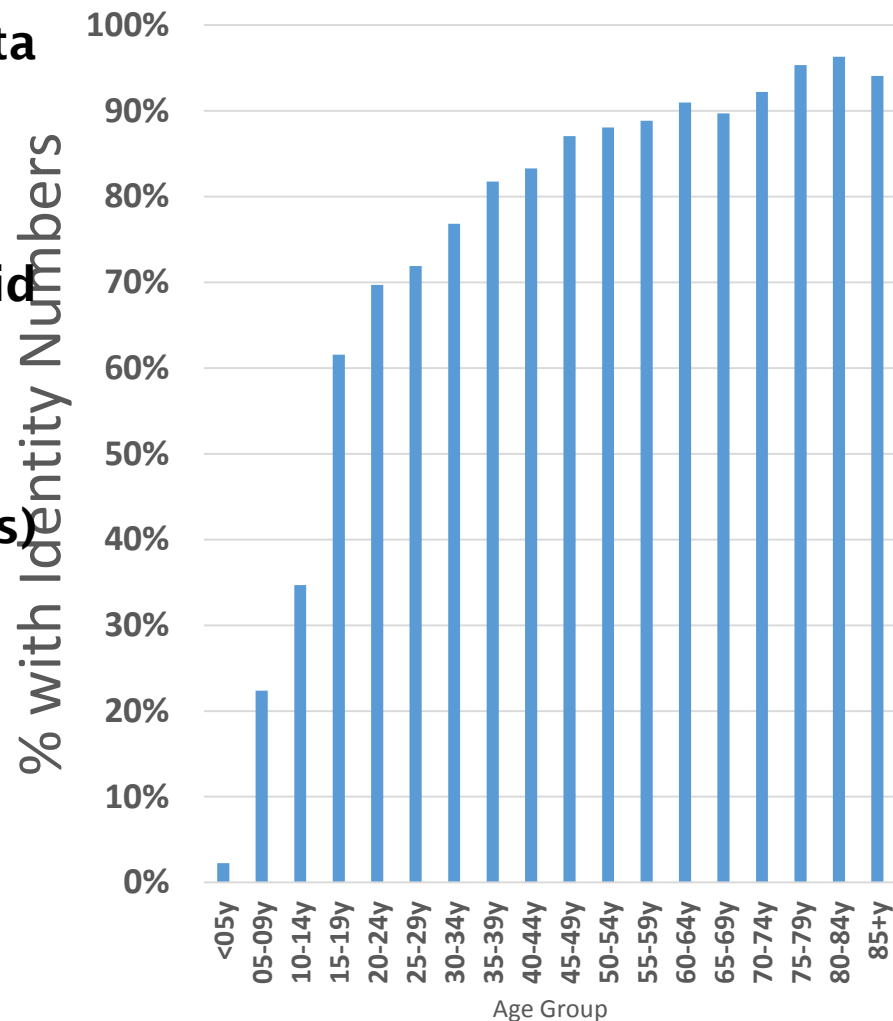
SA Ids in DSA

-81% (n=30,129, ART initiators)

ARTemis

-57% (n=41675) TIER.Net

Proportion of Resident Individuals with Civilian Identity Numbers by Age Group  
1 Jul 2014 (n = 62,213)





# Deterministic data linkage

- Combined variables are used as linkage key

-Individual names+ DOB+sex

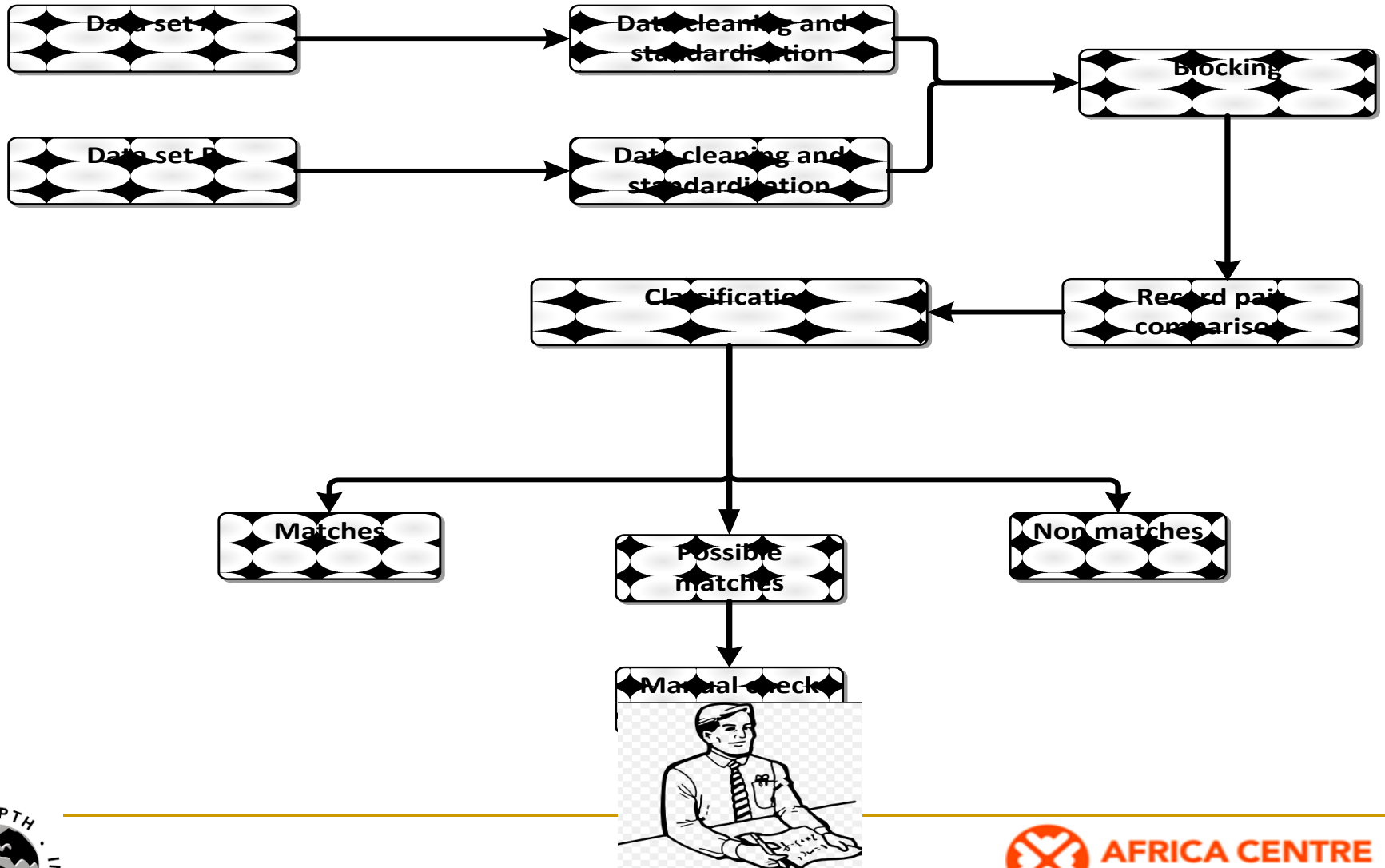
**Validation:** Mother/father names, location and closest clinic

# Probabilistic data linkage

- Common/unique ids rarely available
- Pentaho Data Integration's JaroWinkler Similarity is used
- Weighted score for every possible pair of records
  - Returns jaro similarity coefficient between two strings
  - [SurnameMatch]+[FirstNameMatch]+[SexMatch]+[DoBScore]=Weighted Score
  - A weighted score greater than or equal to a cut off point means match

**Validation:** Mother/father names, location, closest clinic

# Data linkage process



# HDSS data linkage with ART research data

- **HDSS database**

92,000 (62,160 resident) currently

- ~1.3 million person years

10,000 resident adults(age>15) consent for HIV annually

43,307 ever tested, 11,763 positive

- **ART database**

**30,129** initiated ART(Include individuals outside DSA)

Of the HIV+ in HDSS,4,049 (34%) linked to HIV database

# HDSS data linkage with hospital info system data

## Hospital information system (HIS) database

Serves a population=247,350 (2011 SA Census)- (37% from DSA)

-21,881 individuals hospitalized(Include individuals outside DSA)

-2,766 linked to ACDIS

## HIS database with ART database

-4,942 are linked to ARTemis

# Challenges

- Errors, variations and missing data on the info to link the datasets
- Differences in data captured and maintained by different databases
- Data dynamics as data regularly change over time
  - Names and addresses
- Data problems
  - Typos/misspellings (algorithm)
  - Incomplete words
- Methodological issue such as validation of linkages

# Opportunities

- Data linkage helps:

- Improve data quality and integrity

- Reduce costs and efforts in data acquisition

- Allow re-use of existing data sources for new studies

The screenshot shows the Science journal website interface. At the top, there are navigation links for AAAS.ORG, FEEDBACK, HELP, LIBRARIANS, and a search bar. Below this is a red navigation bar with links for NEWS, SCIENCE JOURNALS, CAREERS, MULTIMEDIA, COLLECTIONS, and JOIN / SUBSCRIBE. The main content area displays the article title "High Coverage of ART Associated with Decline in Risk of HIV Acquisition in Rural KwaZulu-Natal, South Africa" by Frank Tanser et al. The article is categorized as a REPORT. The abstract text is visible, starting with "The landmark HIV Prevention Trials Network (HPTN) 052 trial in HIV-discordant couples demonstrated unequivocally that treatment with antiretroviral therapy (ART) substantially lowers the probability of HIV transmission to the HIV-uninfected partner." On the left side, there are several utility links such as "Read Full Text to Comment (0)", "Save to My Folders", and "Download Citation".

## Use of antiretroviral therapy in households and risk of HIV acquisition in rural KwaZulu-Natal, South Africa, 2004–12: a prospective cohort study

Alain Vandormael, Marie-Louise Newell, Till Barnighausen, Frank Tanser

### Summary

**Background** Studies of HIV-serodiscordant couples in stable sexual relationships have provided convincing evidence that antiretroviral therapy can prevent the transmission of HIV. We aimed to quantify the preventive effect of a public-sector HIV treatment and care programme based in a community with poor knowledge and disclosure of HIV status, frequent migration, late marriage, and multiple partnerships. Specifically, we assessed whether an individual's hazard of HIV acquisition was associated with antiretroviral therapy coverage among household members of the opposite sex.

**Methods** In this prospective cohort study, we linked patients' records from a public-sector HIV treatment programme in rural KwaZulu-Natal, South Africa, with population-based HIV surveillance data collected between 2004 and 2012. We used information about coresidence to construct estimates of HIV prevalence and antiretroviral therapy coverage for each household. We then regressed the time to HIV seroconversion for 14 505 individuals, who were HIV-uninfected at baseline and individually followed up over time regarding their HIV status, on opposite-sex household antiretroviral therapy coverage, controlling for household HIV prevalence and a range of other potential confounders.

**Findings** 2037 individual HIV seroconversions were recorded during 54 845 person-years of follow-up. For each increase of ten percentage points in opposite-sex household antiretroviral therapy coverage, the HIV acquisition hazard was reduced by 6% (95% CI 2–9), after controlling for other factors. This effect size translates into large reductions in HIV acquisition hazards when household antiretroviral therapy coverage is substantially increased. For example, an increase of 50 percentage points in household antiretroviral therapy coverage (eg, from 20% to 70%) reduced the hazard of HIV acquisition by 26% (95% CI 9–39).

**Interpretation** Our findings provide further evidence that antiretroviral therapy significantly reduces the risk of onward transmission of HIV in a real-world setting in sub-Saharan Africa. Awareness that antiretroviral therapy can prevent transmission to coresident sexual partners could be a powerful motivator for HIV testing and antiretroviral treatment uptake, retention, and adherence.

**Funding** Wellcome Trust and National Institute of Child Health and Human Development (US National Institutes of Health).

## Articles



**Lancet Glob Health 2014; 2:e109–115**  
Department of Sociology, University of Minnesota, Minneapolis, MN, USA (A Vandormael MD); Faculties of Medicine and Social and Human Sciences, University of Southampton, Southampton, UK (Prof M-L Newell PhD); Africa Centre for Health and Population Studies, University of KwaZulu-Natal, KwaZulu-Natal, South Africa (M-L Newell, T Barnighausen MD, Prof F Tanser PhD); and Department of Global Health and Population, Harvard School of Public Health, Boston, MA, USA (T Barnighausen)

Correspondence to: Alain Vandormael, Department of Sociology, University of Minnesota, Minneapolis, MN 55455, USA  
vandorm@twinstate.edu

## Conclusion

To fill the gaps in the existing data infrastructure, additional steps that integrate HDSS data with other data sources are required to improve data quality

- Effective and efficient work flow to do it routinely

Without such an effort, HDSS are likely to continue to build on fragmented and often costly system with limited access

In the end, the potential for HDSS data will not be fully realized



# Acknowledgement



**INDEPTH Network**  
Better Health Information for Better Health Policy

welcome<sup>trust</sup>



**INDEPTH Network**