The South African Cancer Match Study - A Pilot Study Towards Precision Public Health

Contact: Mazvita Sengayi
National Cancer Registry
Phone: +27118855304
mazvita.sengayi@nhls.ac.za

Mazvita Sengayi[1,2], Wenlong Carl Chen[1,3,4], Adrian Spoerri[5], Elvira Singh[1,2], Matthias Egger[5,6], Julia Bohlius[5] for the IeDEA – Southern Africa Collaboration

[1] National Cancer Registry, National Health Laboratory Service, South Africa, [2]School of Public Health, University of the Witwatersrand, South Africa [3]Sydney Brenner Institute for Molecular Bioscience, University of the Witwatersrand, South Africa,[4]Division of Human Genetics, School of Pathology, University of the Witwatersrand, South Africa, [5]Institute of Social and Preventive Medicine, University of Bern, Switzerland, [6]Centre for Infectious Disease Epidemiology and Research (CIDER), School of Public Health and Family Medicine, University of Cape Town, South Africa

Background and aim

- Precision Public Health capitalizes on large population-based data with the aim to provide targeted interventions for well-defined populations. We used machine-learning techniques to explore the feasibility of using laboratory HIV and cancer data to create a national cohort of HIV-positive people with cancer outcomes in South Africa.

Methods

- We used HIV-ELISA, CD4 cell counts and HIV-RNA loads stored at the central data warehouse of the National Health Laboratory Service (NHLS) for the period 2004-2013 from one rural South African province (Northern Cape Province).
- We used machine-learning (Dedupe[1]) to identify records belonging to the same person in the HIV dataset to create an HIV cohort (linkage within HIV database) and linked it to the National Cancer Registry (NCR, linkage between datasets). Linkage variables included first names, surnames, date of birth and sex.
- We calculated cancer incidence rates per 100,000 person-years from the date of first CD4 count to cancer diagnosis or last CD4 count date. We present cancer incidence rates stratified by sex.

Results

- We identified 653,071 laboratory records and created a cohort of 212,746 unique HIV-positive people. Of these, we analyzed 80,204 patients with ≥2 CD4 count measurements.
- 66% were female, median age was 32 years (IQR 25-40) and median first CD4 cell count 310 cells/µL (IQR 169-493). We identified 1,410 cancers, 889 of whom were incident.
- Overall cancer incidence per 100,000 pys was 518 in men and 367 in women. The most frequent cancers were Kaposi sarcoma, cervical cancer, breast cancer and basal cell carcinoma of the skin, see Figure 1.

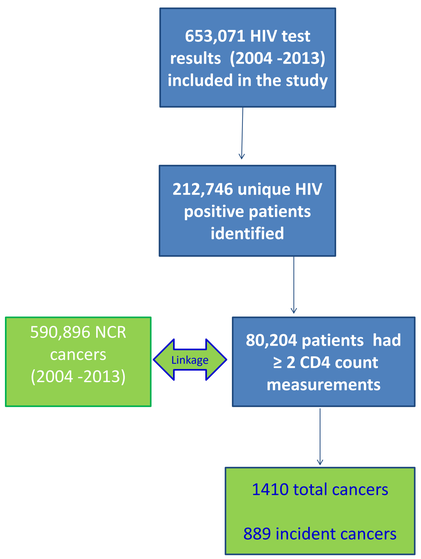
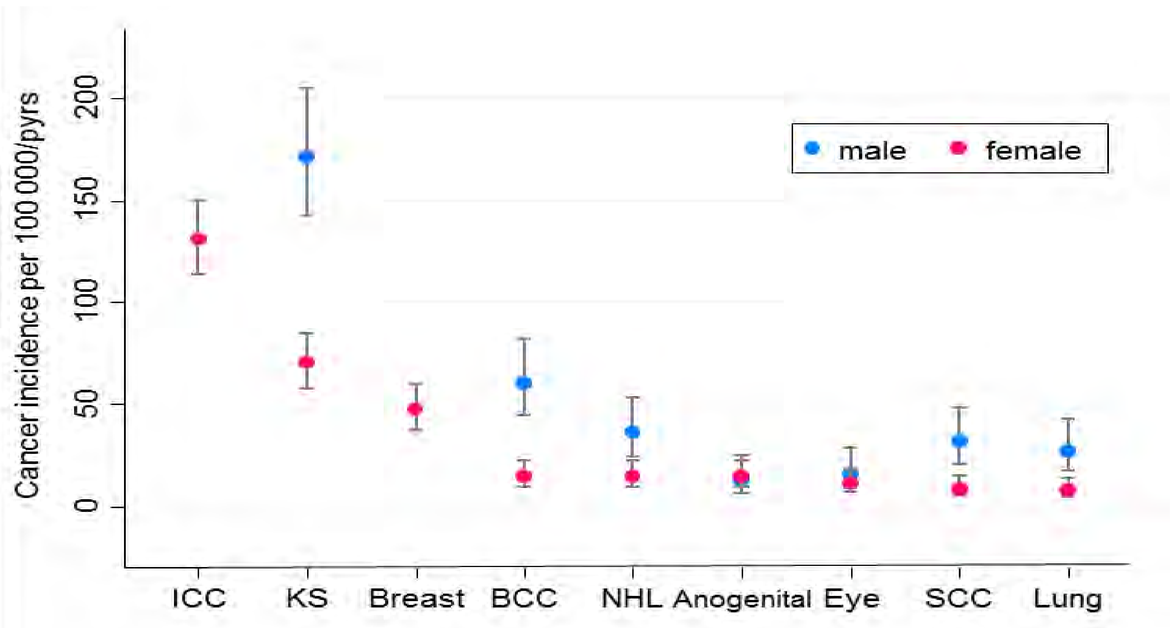


Figure 1: Cancer Incidence in HIV-positive Men and Women in Northern Cape Province, South Africa



Legend: ICC invasive cervical cancer, **KS** Kaposi sarcoma, **BCC** Basal Cell Carcinoma, **NHL** Non Hodgkin Lymphoma, **Anogenital** (vulva, vagina, anus and penis), **SCC** Squamous Cell Carcinoma of the Skin.

Conclusions

- Machine-learning techniques allow creating a national HIV cohort with cancer outcomes based on large laboratory data warehouse and cancer registry data.
- The spectrum of cancers seen in this pilot are similar to those shown in previous studies in South Africa, however, incidence rates are much lower, probably reflecting the poorer access to cancer diagnosis in this rural province.
- The South African HIV Cancer Match study will be used to identify disadvantaged populations and sets the foundation for an era of Precision Public Health.

Funding sources
Research reported in this presentation was supported by the National Institutes of Health administrative supplement to Existing NIH Grants and Cooperative Agreements (Parent Admin Supp) (The South African HIV Cancer Match Study; U01AI069924-09, PI Matthias Egger, co-PI Julia Bohlius) PEPFAR supplement (PI Matthias Egger) and the Swiss National Science Foundation (The South African HIV cancer Match Study, 320030_169967, PI Julia Bohlius) allocated for this project. The contents are solely the responsibility of the authors and do not necessarily reflect the views of the funding bodies.

References
1. https://github.com/dedupeio/dedupe.

CROI 2018 Poster Number 650