International Epidemiology Databases to Evaluate AIDS

General overview and cancer projects involving AI

FG-AI4H – Focus Group AI for Health

PD Dr. Julia Bohlius

Institute of Social and Preventive Medicine,
University of Bern

International Epidemiology Databases to Evaluate AIDS

What is IeDEA?

> International research consortium established in 2005 to provide rich resources for globally diverse HIV/AIDS data

> Funded by U.S. NIH

> Observational HIV care cohorts across 7 geographic regions
  — Each region collaborates with HIV clinical sites to identify, define and collect key variables and implement methodology to analyze data
  — Data are gathered during **routine** clinical (outpatient) HIV care

> Generates large data sets: currently close to 2 million HIV patients enrolled in care and treatment programs

> Data are used to address high priority HIV/AIDS research questions unanswerable by a single cohort

IeDEA 1.7 Mio HIV+ patients, 300+ clinics, 40+ countries

North American AIDS Cohort Collaboration of Research and Design (NA-ACCORD)
- Canada
- United States

West Africa
- Benin
- Burkina Faso
- Côte d'Ivoire
- Ghana
- Guinea-Bissau
- Mali
- Nigeria
- Senegal
- Togo

Asia and Australia
- Australia
- Cambodia
- China
- Hong Kong SAR
- India
- Indonesia
- Japan
- Malaysia
- New Zealand
- Philippines
- Singapore
- South Korea
- Thailand
- Vietnam

Caribbean, Central and South America (CCASAnet)
- Argentina
- Brazil
- Chile
- Haiti
- Honduras
- Mexico
- Peru

Central Africa
- Burundi
- Cameroon
- Rwanda
- Democratic Republic of Congo
- Republic of Congo (RoC)
- Rwanda

East Africa
- Kenya
- Tanzania
- Uganda

Southern Africa
- Lesotho
- Malawi
- Mozambique
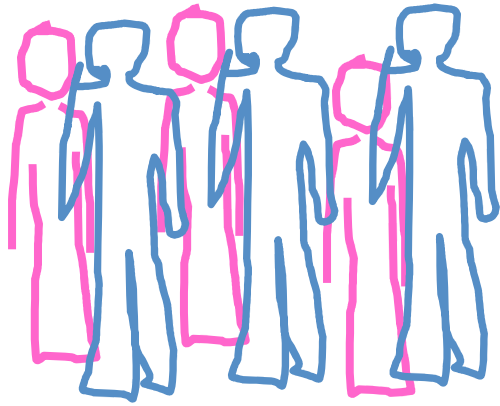- South Africa
- Zambia
- Zimbabwe

IeDEA Southern Africa projects

> Prevention of vertical HIV transmission

> Tracing studies to identify patients who do not return to care

> WHO collaboration to monitor CD4 counts

> ...

> **Cancer studies**
— Big data for Precision Public Health
  – *Burden of disease*
— Evidence based interventions:
  – *Portable colposcope and image recognition*

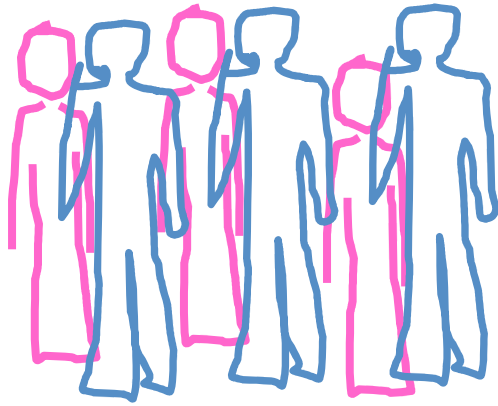# BIG DATA FOR PRECISION PUBLIC HEALTH
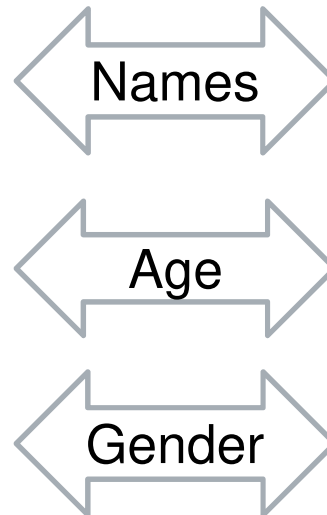
Imperfect data in Southern Africa

HIV cohorts

Probabilistic record linkage in South Africa

Linkage variables

Names

Age

Gender

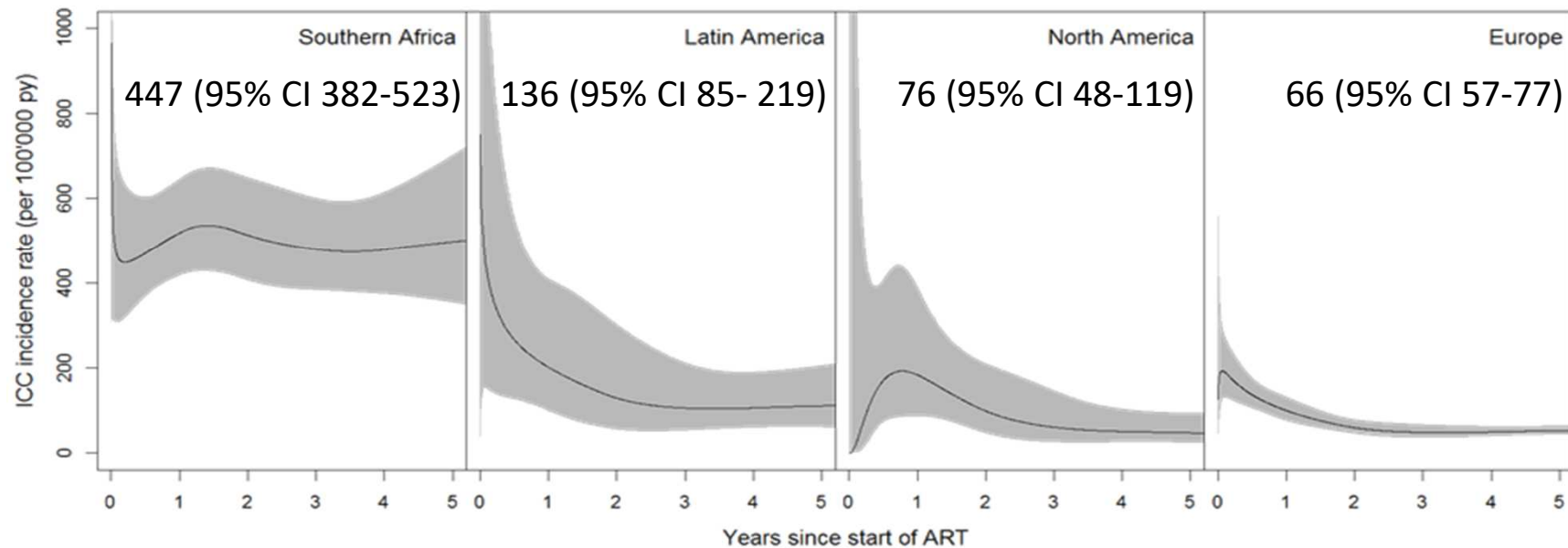**HIV cohorts**   **Cancer registry**

Before…
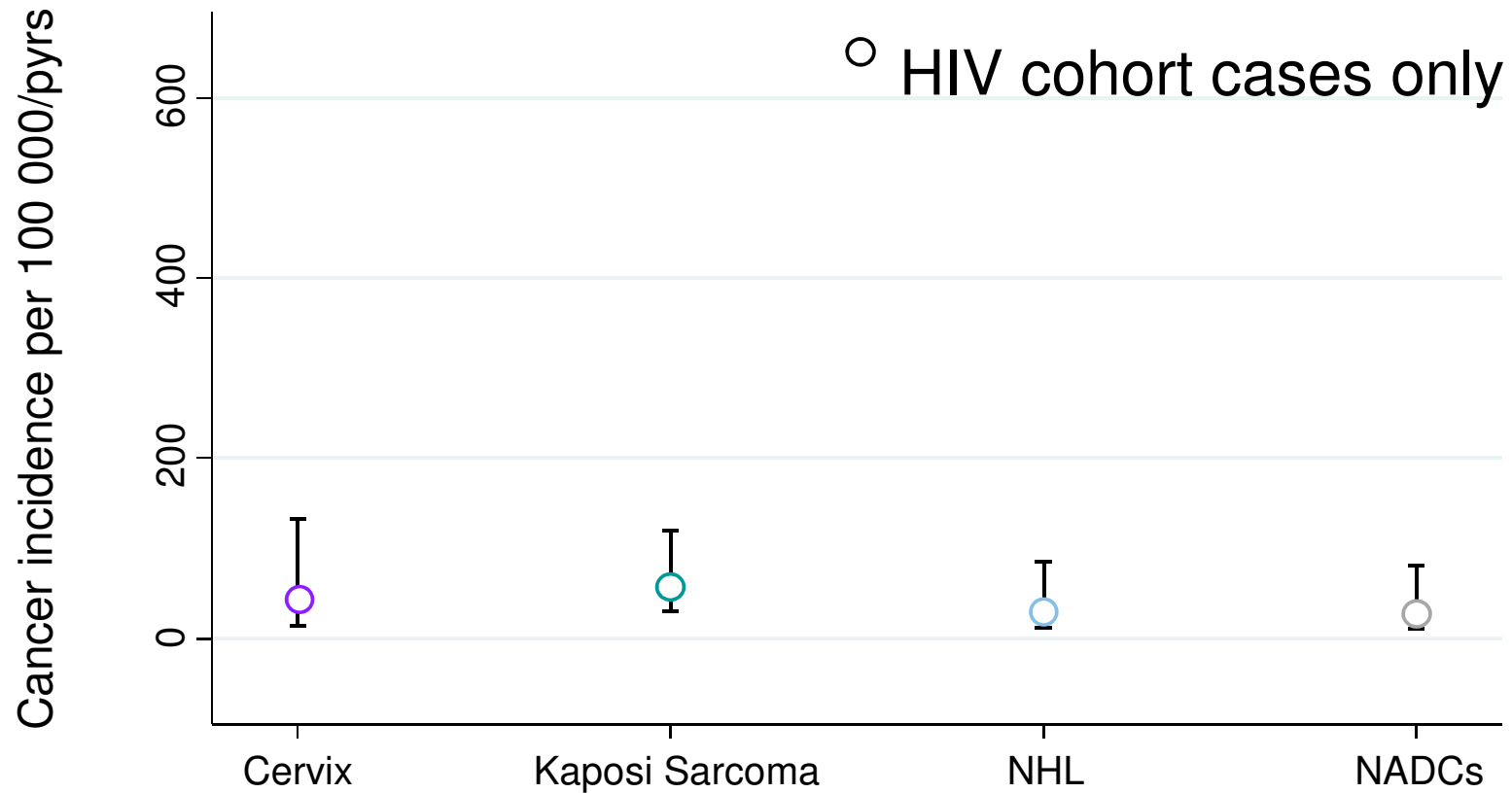
…and after linkage with cancer registry
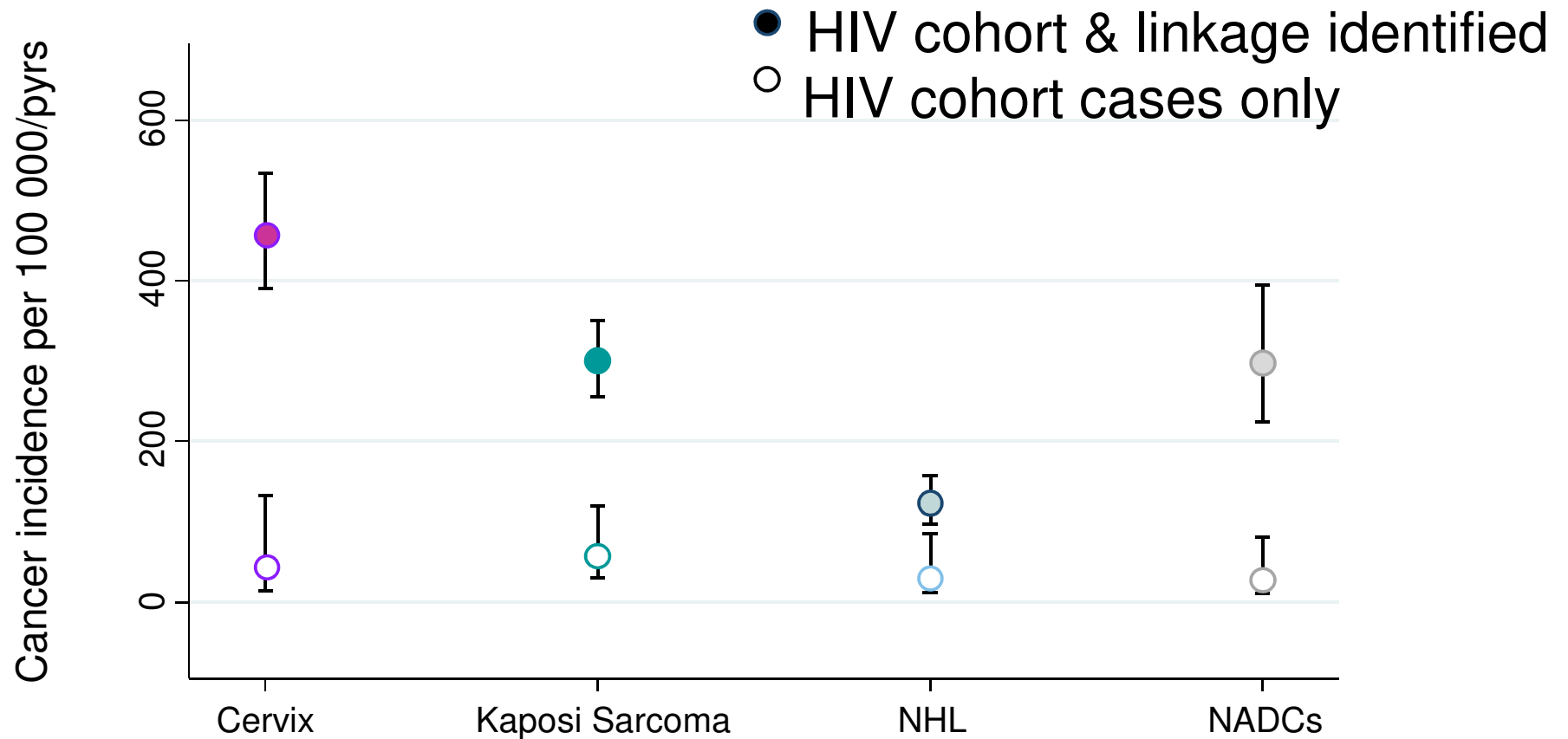
International comparison of cervical cancer risk

> Analyzed 65,726 women, 323,224 person-years

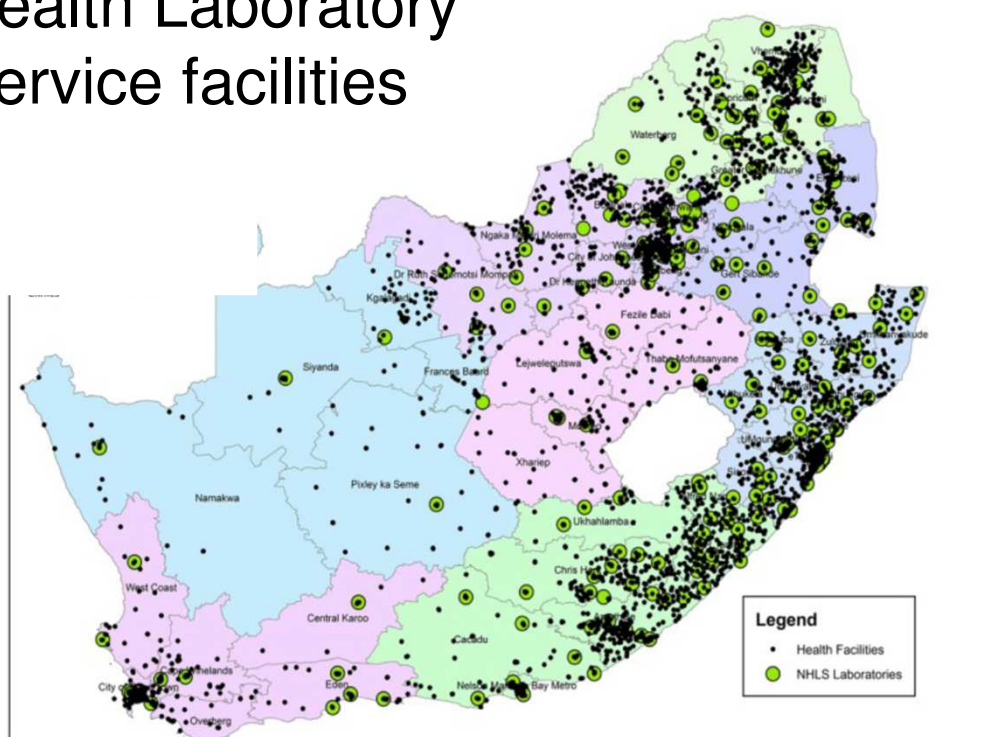> 356 incident cancer cases

> Southern Africa versus Europe: 12.5 (95%CI 7.8-20)*

Rohner et al, under review
Lancet HIV

South African HIV Cancer Match Study

> National Health Laboratory Services founded in 2004

> > 4,000 health facilities, 265 laboratory facilities

Map of National
Health Laboratory
Service facilities

> All laboratory test results are sent to a single data warehouse in Johannesburg

> 50 million records
> > HIV tests
> > CD4 cell counts
> > HIV RNA
> > any other test

> Largely no unique patient identifiers

South African HIV Cancer Match Study

> Methods used

— privacy preserving probabilistic record linkages

— machine learning

> Final data base will include all South African provinces

> 4 – 10 million HIV-positive individuals

> > 100,000 linked cancer cases

PORTABLE COLPOSCOPE AND IMAGE RECOGNITION

Cervical cancer screening methods

> Cervical cancer is a preventable disease
  — Primary prevention: HPV vaccination
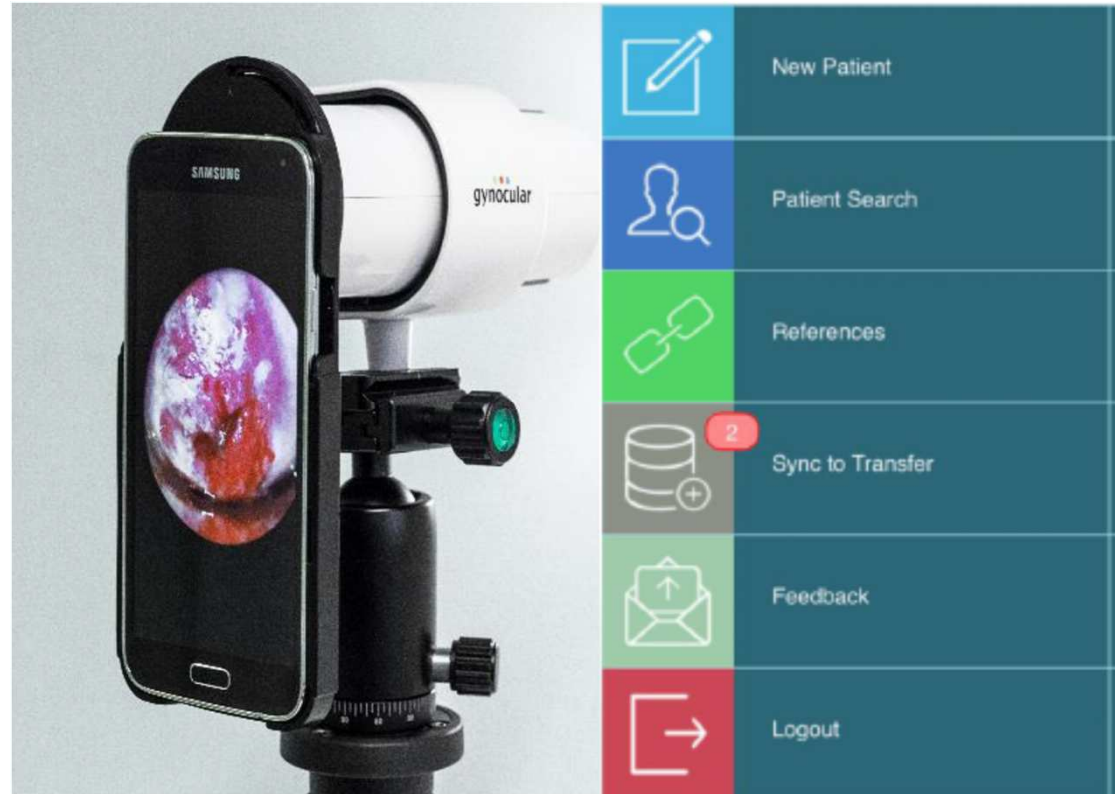 — Secondary prevention: screening for pre-cancerous lesions

	Feasibility	Sensitivity	Specificity
Cytology (Pap)	🔴☹	🟡😐	🟢🙂
Visual Inspection with Acetic Acid	🟢🙂	🔴☹	🟡😐
HPV testing	🟡😐	🟢🙂	🔴☹
Portable colposcope	🟢🙂	?	?

Portable colposcope

> Magnification

> Smartphone

> Images

> Telemedicine

Clinical trial starting in Lusaka in 2018

> 450 women

— VIA

— HPV

— Portable colposcope

> Histopathology

Funder: Swiss Cancer Research

Visual inspection of the cervix

> **Cervix Image Sharing Protocol (CISP)**
  — NCI project, Principal Investigator: Julia C Gage, Ph.D.
  — https://clinicaltrials.gov/ct2/show/NCT03179306

> **Cervical images from four NCI studies**
  — Costa Rican Natural History Study of HPV and Cervical Neoplasia
  — ASCUS LSIL Triage Study-- ALTS
  — SUCCEED
  — Costa Rica Vaccine Trial

> **78'787 images, CIN2+ 13'073 patients (10.4%)**

> **Algorithms to detect CIN2+ currently being developed**

Needs identified in both projects

> IT infrastructure

> Skilled work force and capacity building

> Funding

> Reference data sets to validate algorithms

> Reporting guidelines

Acknowledgements – IeDEA-SA

Site investigators and cohorts:
Gary Maartens, Aid for AIDS, South Africa; Michael Vinikoor, Centre for Infectious Disease Research in Zambia (CIDRZ), Zambia; Monique van Lettow, Dignitas, Malawi; Robin Wood, Gugulethu ART Programme, South Africa; Nosisa Sipambo, Harriet Shezi Clinic, South Africa; Frank Tanser, Africa Centre for Health & Population Studies (Hlabisa), South Africa; Jonathan Euvrard, Khayelitsha ART Programme, South Africa; Geoffrey Fatti, Kheth'Impilo, South Africa; Sam Phiri, Lighthouse Clinic, Malawi; Cleophas Chimbetete, Newlands Clinic, Zimbabwe; Karl Technau, Rahima Moosa Mother and Child Hospital, South Africa; Brian Eley, Red Cross Children's Hospital, South Africa; Josephine Muhairwe, SolidarMed Lesotho; Anna Jores, SolidarMed Mozambique; Kamelia Kamenova, SolidarMed Zimbabwe, Matthew P Fox, Themba Lethu Clinic, South Africa; Hans Prozesky, Tygerberg Academic Hospital, South Africa.

Data centers:
Nina Anderegg, Marie Ballif, Julia Bohlius, Cam Ha Dao Ostinelli, Frédérique Chammartin, Matthias Egger, Lukas Fenner, Per von Groote, Andreas Haas, Taghavi Katayoun,Eliane Rohner, Adrian Spörri, Gilles Wandeler, Elizabeth Zaniewski, Kathrin Zürcher, Institute of Social and Preventive Medicine, University of Bern, Switzerland; Andrew Boulle, Morna Cornell, Mary-Ann Davies, Victoria Iyun, Nicola Maxwell, Leigh Johnson, Patience Nyakato, Gem Patten, Michael Schomaker, Priscilla Tsondai, School of Public Health and Family Medicine, University of Cape Town, South Africa.

Lab of Digital Epidemiology

Sharada Prasanna Mohanty, PhD Student
Professor Marcel Salathé
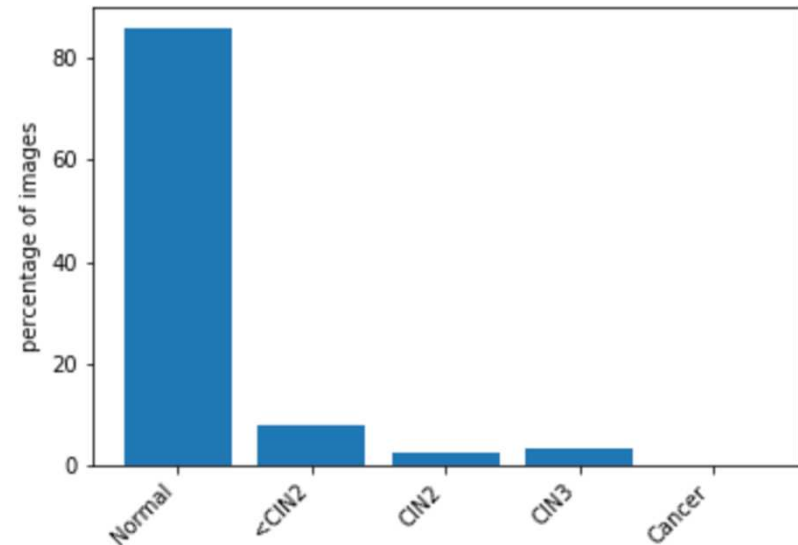Camille Montalcini, PhD Student
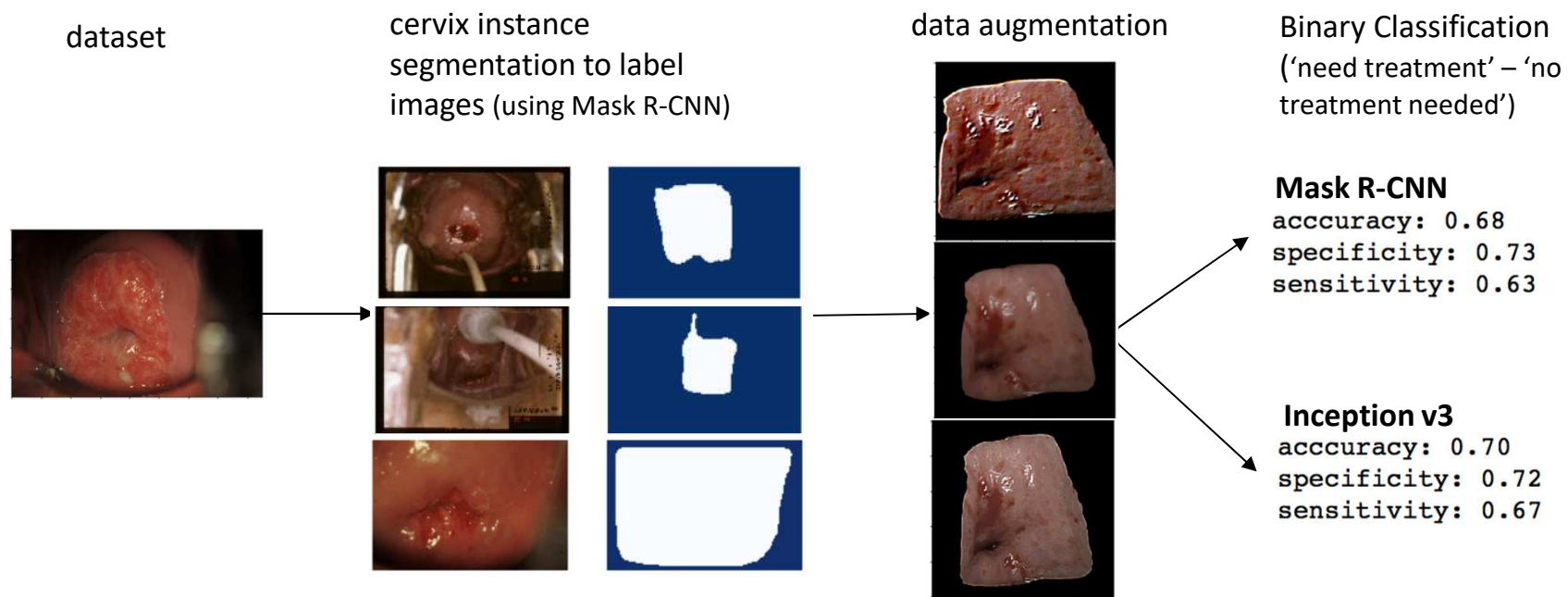
Steps to creating the basic algorithm

> Identify the cervix
>> – Used R-CNN
>> – Manually marked out 60 cervices
>> – Identification of the "Mask"

> Is treatment required or not
>> – Used Inception V3
>> – Images augmented
>> – Can add more layers
>> to make the capacity of the
>> algorithm more complex too, later

> Testing on a balanced database

Process to predict when treatment is needed (stage CN2 or higher)

dataset

cervix instance segmentation to label images (using Mask R-CNN)

data augmentation

Binary Classification ('need treatment' – 'no treatment needed')

Mask R-CNN
acccuracy: 0.68
specificity: 0.73
sensitivity: 0.63

Inception v3
acccuracy: 0.70
specificity: 0.72
sensitivity: 0.67

Future directions

Infinite learning?!

IeDEA Southern Africa Steering Group Meeting, Johannesburg, South Africa 2017