

Talk

# The Evolution of Data Analytics in Healthcare

Africa Digital Epidemiology Innovation Network  
Speaker Series  
May 2022



**Dorcas Kareithi, Msc**

Associate Statistician (Clinical Trials)  
Newcastle University  
Biostatistics Research Group.



**Steven Karera, Msc**

Data Analysis Associate  
CCHUB Design Lab.

# 1

## Background



Over 8 years of experience as a statistician, researcher and analyst. Worked as an analyst, head of research and consulting epidemiological analyst in several organizations and companies over the years. My research interests are in oncology trials and statistical computation.

Mentor, current chairperson of Young African Statisticians (YAS) and member of IYASA-KE, IBS and ISI.

<https://www.linkedin.com/in/dorcaskareithi/>

The Population Health Sciences Institute Biostatistics Research Group (PHSI-BRG) at Newcastle University consists of 25 applied statisticians (and growing) with a variety of backgrounds, experience, and expertise.

The PHSI-BRG provides statistical expertise to a large portfolio of clinical trials and observational studies, working closely with Newcastle Clinical Trials Unit and the NIHR Research Design Service North East and North Cumbria.

As well as supporting real clinical trials, the PHSI-BRG has a newly established and growing methodology research group.

<https://www.newcastle-biostatistics.com/>

**LET'S TALK  
ABOUT DATA,  
DISEASES, &  
DECISIONS.**



**HEALTHCARE**





The process of **producing, processing and consuming** data from and about the **health of an individual or collective population**. This information is from **Health Information systems, research projects or other technological tools** utilized by people in the healthcare ecosystem.

# 4

## Health Data Analytics Pipeline

### DATA SOURCES

Obtaining and extracting data from different sources

### DATA CONSUMPTION

Translating the insights to policy and practice. Data literacy



### DATA PROCESSING

Cleaning, analyzing and “torturing the data”.

### DATA PRESENTATION

Helping audiences understand what you did and the insights you pulled from the data.

# 5

## Evolution of Sources of Data

- Traditionally, we had a limited set of sources of data- specifically in healthcare
- The rise of the digital era, more and more sources of data become available
- The healthcare industry has seen a wider range of sources of data
- Ensuring this data is of great quality, reliable and verifiable is still evolving, with great steps made in that direction

- Patient paper records
- Civil Registration & Vital statistics
- Population census records
- Research studies data
- Clinical trials data
- National Statistical Offices
- Community sources of data
- Household and Demographic health surveys

- Medical Images
- GIS/Map data
- Electronic Prescription Services (E-prescribing)
- Social media data
- Electronic health management data
- Data from wearables
- Historical data
- Third-party healthcare data (insurance, other software)



# 6

## Evolution of Data Processing / Analysis

- **Descriptive Analytics:** Descriptive analytics uses historical data to draw comparisons or discover patterns. This type of analysis is best for answering questions about what has already occurred
- **Predictive Analytics/Modelling:** Predictive analytics uses current and historical data to make predictions about the future.
- **Multiple simulations forecasting:** People create models to simulate situations, and optimize for the best delivery configuration, such as vaccination priority if supply is limited.
- **Machine learning and AI.** Artificial intelligence modules to help detect abnormalities, early outbreaks and monitor the mobility and evolution of diseases via timely data analytics.
- **Personalized/precision medicine and advanced genetic research.** Biomedical research has seen the greatest evolution over the years. Using vital genetic information and other factors to recommend treatment
- **Adaptive clinical trials.** evaluates a medical device or treatment by observing participant outcomes on a prescribed schedule, and, uniquely, modifying parameters of the trial protocol in accord with those observations.
- **Use of statistical programming languages, data science macros and packages.** Increase in popularity in use of open source software like R, Python and inbuilt macros and packages.



Tables and  
charts



Dashboards



Newsletters &  
Newspapers

**TRADITIONAL  
METHODS**



**MODERN  
METHODS**

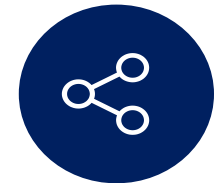
**Methods complement people in all  
sides of the digital divide.**



Personalized  
medical reports



Infographics &  
Videographics



Social Media Posts





- **Data for decision making/policy.** Triggered by outbreaks. Many examples of how governments and policy makers are quick to embrace data for decisions post outbreaks.
- **Data to empower healthcare professionals.** Decisions made by health professionals are edging towards evidence-based medicine.
- **Data to improve services.** Queues can be shortened via quicker service. Lives can be saved as well with information quickly accessed in one central point.
- **Tele-health services.** People are able to access the best of healthcare from the comfort of their home. From diagnosis to prescription and patient history
- **Health data literacy for the population.** There has been a shift in people being concerned about their health and some develop good healthcare seeking behavior. This is still a work in progress in Africa
- **Misinformation in healthcare.** Increase in information means increase in misinformation. All generations have experienced this. We need joint effort from everyone in the healthcare ecosystem.

- **Limited use.** Data generated in the routine care of patients may be limited in its use for analytical purposes.
- **Inaccurate or Incomplete data.** A lot of data from African countries, even in continental databases is either missing or outdated.
- **Causality/inaccurate reporting.** Clinical data mostly only allows observational and not experimental studies, thus raising issues of cause-and-effect of findings discovered.
- **Data driven research.** Some researchers note that research questions asked of the data tend to be driven by what can be answered, as opposed to prospective hypotheses
- **Data wastage.** This mostly comes in as a result of collection of excessive data. We've become very efficient at getting data into a repository and structuring it—but gathering insights and putting them to use is still challenging.
- **Ethical challenges:** who owns data and who has privileges to use. This is exemplified in the digital era and with the evolution of data analytics in healthcare
- **Lack of standardization.** Data may also incompletely adhere to well-known standards, which makes combining it from different sources more difficult
- **Duplication of effort (interoperability challenges).** Multiple people/institutions collecting the same data in the same area using the same population/sample. Could open data solve this?

“Diverse (in complexity and type) data generated from electronic health records that they cannot be processed with traditional hardware, software or conventional data management tools and methods. This includes, clinical data, doctor's comments and prescriptions, medical imaging, laboratory test results, pharmaceutical records, health insurance reports, administrative reports, patient historical data, posts on social media platforms such as, Facebook and Twitter, blogs, and other information that are not directly connected to patients, such as, news in health magazines and relevant publications in medical journals”  
(Raghupathi & Raghupathi, 2014).

## Why?

- Need to continuously improve patient health outcomes
- Shift in evidence-based medicine
- Continue to improve the efficiency of clinical processes and other healthcare operations
- Making healthcare more affordable
- Enhance the diagnosis and the prescription of suitable treatment pathways for patients
- Ensure healthcare systems monitoring and augmented accountability

## How can we leverage/contribute?

- Building infrastructure that can help advance use big data in Africa such as cloud-based databases, repositories and African-based algorithms
- Train more healthcare personnel and build capacity in data collection, processing and presentation
- Work on legislation to protect patient data by involving law and policy makers
- Highlight and appreciate healthcare systems already making progress
- Is open health data an option?

“Open data is large shared medical datasets and data-driven research. Open data can shed light on the causes of disease and effects of treatment, including adverse reactions side-effects of treatments, while also facilitating analyses tailored to an individual's characteristics, known as personalized or “stratified medicine.” Developments, such as crowdsourcing, participatory surveillance, and individuals pledging to become “data donors” and the “quantified self” movement (where citizens share data through mobile device-connected technologies), have great potential to contribute to our knowledge of disease, improving diagnostics, and delivery of healthcare and treatment.” Kostkova et al (2016)

- We need a centralized, systematic way of collecting, storing and analyzing data so we can use it to our advantage
- There are a lot of obstacles when it comes to sharing health data across organizations and a distinct lack of standardization in the way that data is collected and analyzed.
- Major concerns over privacy, confidentiality, and control of data about individuals once it is shared
- Growing examples of open data in Africa on Google:
  - [Ehealthafrica Data Portal](#)
  - [AfricaOpenData](#)
  - [Data Africa](#)
  - [AfDB Africa Health Atlas](#)
  - [NIH-Harnessing Data Science for Health Discovery and Innovation in Africa \(DS-I Africa\)](#)
  - Global platforms from WHO, Meta, HDX, CDC

**5 min presentation by each group**

1. **Have you encountered open and big data in your day to day activities**
  - What data was/is it?
  - How did you use it?
2. **Open data in healthcare:**
  - Is it important?
  - What are the advantages?
  - What are the disadvantages?
3. **Would you be open to sharing data to the network?**

**DAY**



# **Practical Session**





STEVEN KARERA



**Steven Karera, MSc**, is an analyst with over 6 years' experience implementing and assessing projects across various sectors. . He has strong knowledge in statistical methods and data analysis. He is currently working as data and analytics lead at Co-creation Hub, design lab and he holds a master's degree in data science.

Mr. Steven is a skilled analyst with strong knowledge in data modeling and management of unconventional datasets. He mainly works on health data, and social network data while focusing on statistical methodologies, estimating inference, predictive and prescriptive modeling, and data visualizations.

<https://www.linkedin.com/in/steven-karera-50719310b>

The CcHUB Design Lab is the Research and Development Centre with a mandate to:

- Incorporate Human-Centered Design (HCD) into how public agencies and corporates (Medium to large) solve problems
- Exploit technology to build and improve services, product and process
- Build partnerships and relationships to collaboratively solve significant social and business challenges across Africa

The lab is powered by a multidisciplinary design team that includes designers, engineers and researchers who collaborate to explore the application of emerging technologies that will solve Africa's systemic problems in Public Health, Education, Governance and the Private Sector.

<https://cchubnigeria.com/designlab>

- Open data: data that can be freely used, re-used and redistributed by anyone.
- Open data may lead to more accurate conclusions and better decisions.
- Generates new knowledge from combined data sources and patterns in large data volumes
- Beneficiaries: citizens, government and business environment.



The Humanitarian Data Exchange (HDX): an open platform for sharing data. Launched in 2014, aiming to make humanitarian data easy to find and use for analysis. The platform has been accessed by users in over 200 countries and territories.



WHO's Global Health Observatory (GHO): Contains health-related statistics for almost 200 Member States. The over 1,000 indicators collected are mainly used to monitor progress towards the health aspects of the Sustainable Development Goals.

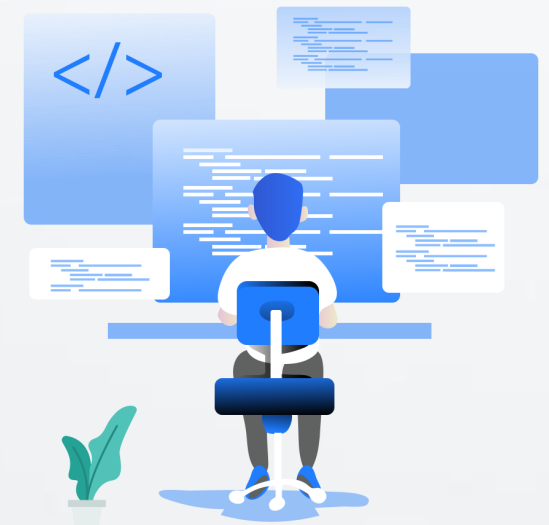
- R is a **freely available, open source** “computational language and environment for data analysis and graphics.”
- Reproducible scripts, functions and packages
- As researchers, epidemiologists, data analysts, and health practitioners, we use R in the following ways:
  - As a full-function calculator
  - To perform analysis and build models
  - To create extensible statistical packages and macros
  - To build dashboards to communicate findings
  - To build presentations for conferences and meetings
  - To create real-time applications that can be embedded on platforms



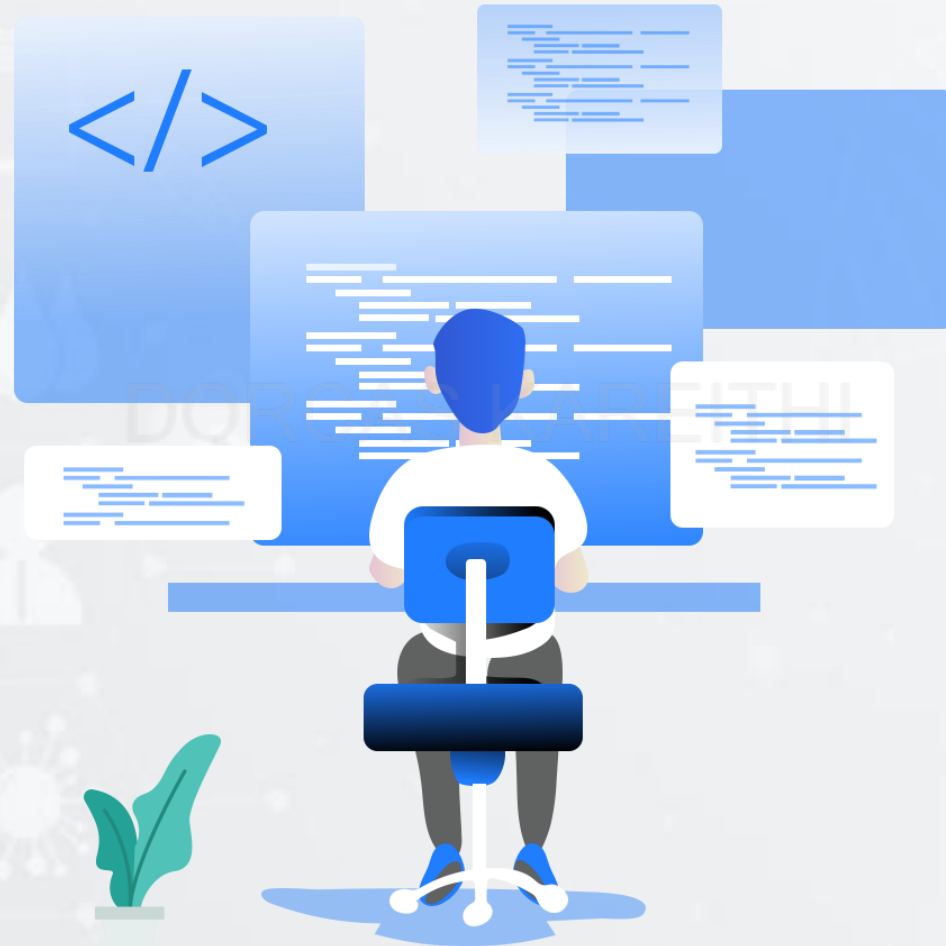
(This list and instructions will be shared with you)

See how to install packages here: [https://www.youtube.com/watch?v=LUhU\\_GX6fys](https://www.youtube.com/watch?v=LUhU_GX6fys)

- **Dplyr**: for data manipulation
- **Tidyverse**: an opinionated collection of R packages designed for data science
- **Countrycode**: helps to handle country names and country codes
- **Devtools**: This will help you to install rgho. Ensure you have updated R or
- **Rgho**: Accessing WHO Global Health Observatory – you may need to use  
“devtools::install\_github("pierucci/rgho@devel") “
- **funModeling**: Exploratory data analysis and data preparation tool
- **Plotly**: for creating interactive graphs
- **Sjlabelled**: Collection of functions dealing with labelled data
- **Flexdashboard**: helps to make it easy to create interactive dashboards for R, using R Markdown
- **Shiny**: package that makes it easy to build interactive web apps straight from R
- **Htmltools**: package that makes it easy to customize the user interface (UI) of any Shiny or R Markdown project by using R code.



# LET'S GET CODING!



# RESOURCES PLUG



- ADIEN Platform – [JOIN US](#)
- Resources on using R
  - [R studio Education](#)
  - [Statistical Analysis with R for Public Health](#)
- Resources to learn about what is going on in Public Health and Analytics in Africa
  - [Data Science Nigeria](#)
  - [Utano Public Health Chats](#)
  - [Zindi Africa](#)
  - [Accelerating Artificial Intelligence In Africa](#)



# REFERENCES

- The importance of healthcare data analytics: <https://www.ibm.com/eg-en/watson-health/learn/healthcare-data-analytics>
- Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. Health information science and systems, 2(1), 3. <https://doi.org/10.1186/2047-2501-2-3>
- Kostkova, P., Brewer, H., de Lusignan, S., Fottrell, E., Goldacre, B., Hart, G., Koczan, P., Knight, P., Marsolier, C., McKendry, R. A., Ross, E., Sasse, A., Sullivan, R., Chaytor, S., Stevenson, O., Velho, R., & Tooke, J. (2016). Who Owns the Data? Open Data for Healthcare. Frontiers in public health, 4, 7. <https://doi.org/10.3389/fpubh.2016.00007>
- Mbondji, P. E., Kebede, D., Soumbey-Alley, E. W., Zielinski, C., Kouvidila, W., & Lusamba-Dikassa, P. S. (2014). Health information systems in Africa: descriptive analysis of data sources, information products and health statistics. *Journal of the Royal Society of Medicine*, 107(1\_suppl), 34-45.



# **Asante Sana! Urakoze Cyane!**

**Dorcas Kareithi, Msc**

Associate Statistician (Clinical Trials)  
Newcastle University – Biostatistics Research  
Group.

<https://www.linkedin.com/in/dorcaskareithi/>

**Steven Karera, Msc**

Data and Analytics Lead  
CcHUB Design Lab.

<https://www.linkedin.com/in/steven-karera-50719310b/>