

# **Explorations of DDI for describing data harmonisation in ALPHA network**

Chifundo Kanjala, Tito Castillo, Gareth Knight, David Beckles, Jim Todd,  
Basia Zaba

**$\alpha$**  – network

Analysing  
Longitudinal  
Population-based  
HIV/AIDS data on  
Africa

**Improving health worldwide**

**[www.lshtm.ac.uk](http://www.lshtm.ac.uk)**

LONDON  
SCHOOL of  
HYGIENE  
& TROPICAL  
MEDICINE



# Overview

- Background
- Motivation
- Data harmonisation and our DDI explorations
- Discussion



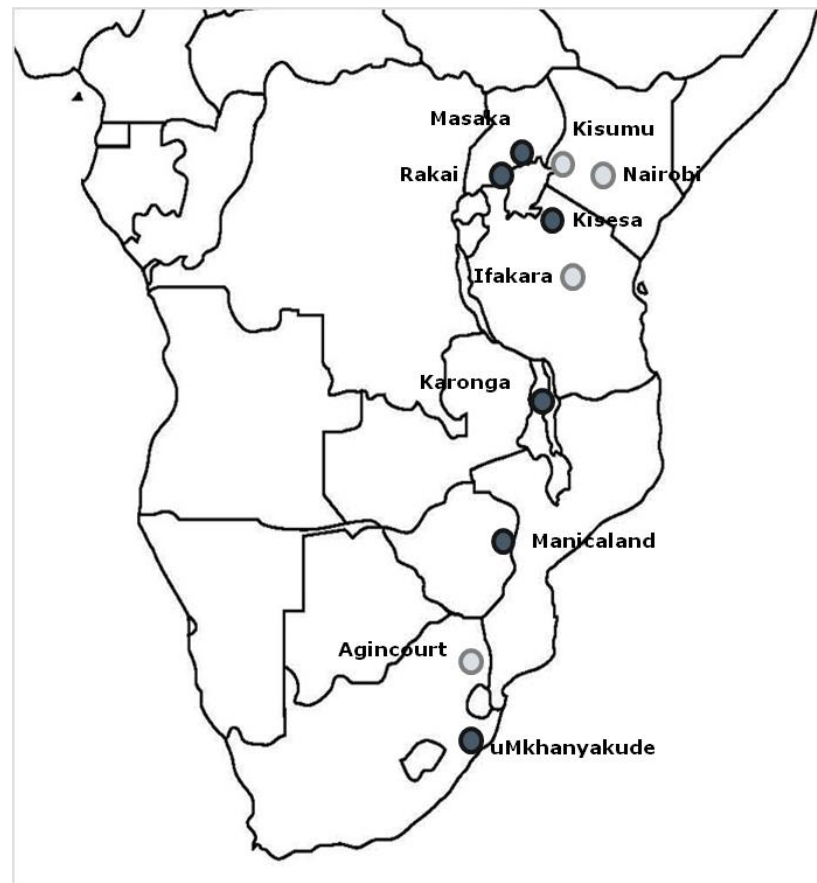


# Background



# ALPHA Network

- A network for Analysing Longitudinal Population HIV/ AIDS data on Africa (ALPHA)
- Runs training workshops
- Analysis of demographic impacts of HIV infection
- 10 autonomous research centres
- Centres contribute data in uniform Stata format



# ALPHA broad analytical themes

- Age patterns of HIV infection
- HIV and fertility
- Mortality in the era of national ART programs
- HIV as a determinant of child mobility
- Partnership dynamics and sexual behaviour
  - Biological evidence insufficient for policy guidance
  - Design and evaluation of intervention programs need behavioural data



# Motivation



## **Primary motivation**

- Need to share data beyond ALPHA

## **Other drivers**

- Efficient multi-site data management and exchange among ALPHA members
- Standardisation will improve the use of existing data
- Funders' policies requiring data sharing
- Pooling and sharing data brings credit through data citation.



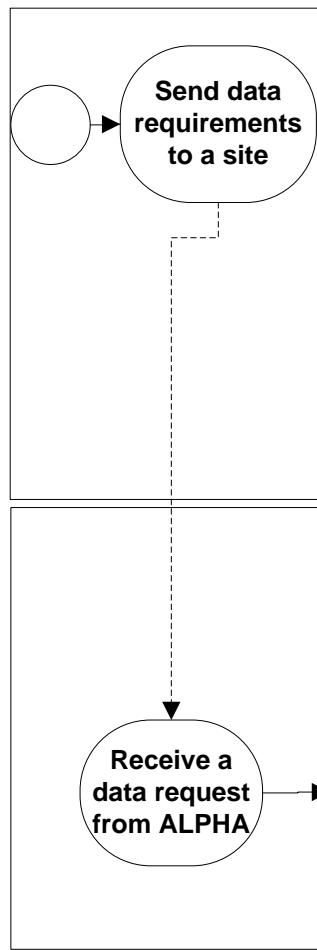


# **ALPHA data harmonisation, DDI use and exploration**





# Preparation of ALPHA datasets



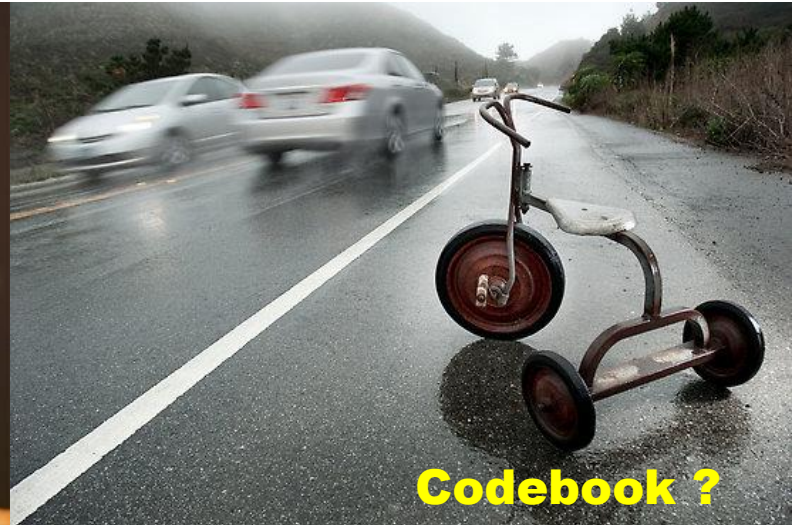
## 6.1 Essential data for each residence episode – one record per episode, including

Variable name	Description	Coding	Notes
idno	Person ID number	site specific	Numeric IDs
study_name	Name of your study field site	site specific	Character –
sex	Male or female	1 Male 2 Female	Must not var
dob	Date of birth- best estimate	in Stata format (days since 1 <sup>st</sup> Jan 1960).	If actual mor impute, e.g. Must not var
residence	Type of area within DSS	site-specific grouping, we expect most sites to have 2 to 4 categories	Aim to distin distinguish r
entry_date	Date of start of residence episode	in Stata format	This date sh the date of a consecutive
entry_type	Type of entry	1 baseline recruitment 2 birth 3 in-migration	

# DDI Harmonisation Descriptions

- Studies use either DDI codebook (Nesstar) or no DDI
- We have DDI codebook descriptions of the ALPHA templates
- Is DDI codebook adequate?

**Lifecycle ?**



# What does DDI Lifecycle offer?

- Group and Comparison modules, Ionescu et al (2009) explored this in detail  
[http://www.ddialliance.org/sites/default/files/UsingDDI3ForComparison\\_0.pdf](http://www.ddialliance.org/sites/default/files/UsingDDI3ForComparison_0.pdf)
  - Group module confirms similarities using inheritance
  - Comparison module shows what the commonalities and differences are between source and target datasets by mapping their concepts, universes, questions, variables, categories and code schemes
- ConceptVariable and RepresentedVariable
  - Found within the LogicalProduct module of DDI Lifecycle
  - RepresentedVariable has the core reusable components of a variable (concept, universe and value domain)



# Sexual behaviour standard variables

- Age at first sex (in completed years)
- On what date did you start having sexual relations with this partner? (DDMMYYYY)
- When was the last date on which you had sex with this partner? (DDMMYYYY)
- How often did you use condoms with this partner? (Never, Sometimes, Mostly, Always)



# Survey questions from 3 ALPHA centres

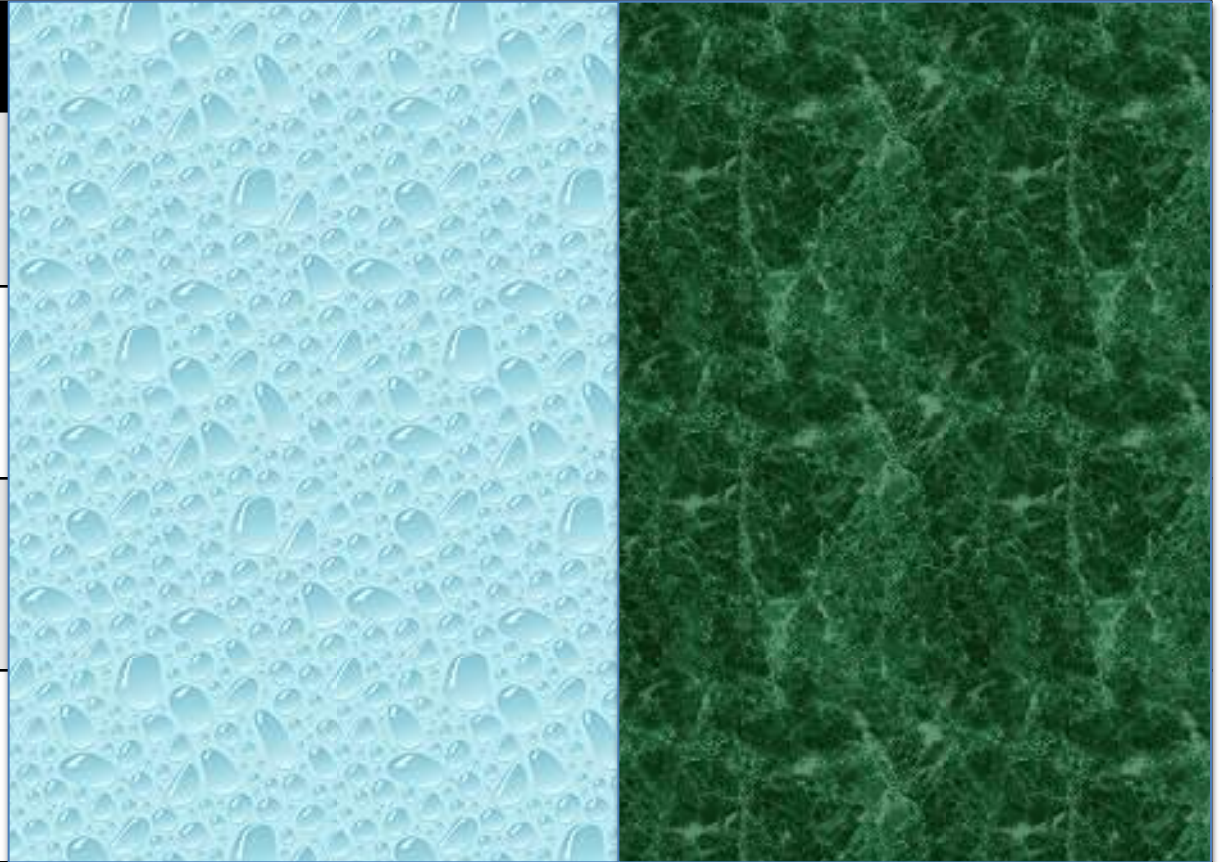
## SA study

How old were you when you first had sex

How long were you sexually involved with this partner?

When was the last time you had sex with partner?

How often did you use condoms with this partner





# Exploring DDI RepresentedVariable

## SA Study

How long were you sexually involved with this partner?  
When was the last time you had sex with partner?

Derived variable  
(Date of first sex with partner)

## Zimbabwe Study

Month and year of first sex with this partner

Date of first sex with partner

RepresentedVariable

## Malawi Study

When was the first time you had sex with this partner? (days, weeks, months or years)

Derived variable  
(Date of first sex with partner)



# Discussion

- What aspects of our data harmonisation can DDI Lifecycle describe?
  - Comparison of source and target datasets (group module, comparison module, RepresentedVariable)
- What DDI tools should we consider?
  - We are looking into Questasy, Colectica, Sledgehammer
  - The tool needs to capture comparison of source and target datasets
  - Describing data transformations in a standardised way (preferably not statistical, SQL or any code)
- What are the aspects that DDI does not describe?
  - Data transformations
- What alternatives to DDI should we look at?
  - Business Process Model Notation (BPMN), SDMX



# Acknowledgements

- ALPHA study centres
- Colleagues at ALPHA secretariat
- Unsung heroes, the study participants
- Funders Wellcome Trust

## Thank you!!!

