Explorations of DDI for describing data harmonisation in ALPHA network

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

Improving health worldwide

www.lshtm.ac.uk
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**

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Description</th>
<th>Coding</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>idno</td>
<td>Person ID number</td>
<td>site specific</td>
<td>Numeric IDs 7-10 digits</td>
</tr>
<tr>
<td>study_name</td>
<td>Name of your study field site</td>
<td>site specific</td>
<td>Character – first 15 characters</td>
</tr>
<tr>
<td>sex</td>
<td>Male or female</td>
<td>1 Male</td>
<td>Must not vary</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 Female</td>
<td></td>
</tr>
<tr>
<td>dob</td>
<td>Date of birth – best estimate</td>
<td>in Stata format (days since 1st Jan 1960)</td>
<td>If actual more than 9999 days, impute, e.g. 10000. Must not vary</td>
</tr>
<tr>
<td>residence</td>
<td>Type of area within DSS</td>
<td>site-specific grouping, we expect most sites to have 2 to 4 categories</td>
<td>Aim to distinguish rural, village, urban, peri-urban</td>
</tr>
<tr>
<td>entry_date</td>
<td>Date of start of residence episode</td>
<td>in Stata format</td>
<td>This date shows the date of arrival in consecutive visits</td>
</tr>
<tr>
<td>entry_type</td>
<td>Type of entry</td>
<td>1 baseline recruitment</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 birth</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 in-migration</td>
<td></td>
</tr>
</tbody>
</table>

Send data requirements to a site

Receive a data request from ALPHA

Receive queries from ALPHA secretariat

Receive data absence notification

Assess quality

Good quality

Further edits

Data available?

Send notification to ALPHA

Yes

Receive site specific ALPHA dataset

No

Send queries to site

Yes

No

ALPHA affiliated research centre

ALPHA network secretariat

Pooled ALPHA dataset
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?
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)
<table>
<thead>
<tr>
<th>Survey questions from 3 ALPHA centres</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SA study</strong></td>
</tr>
<tr>
<td>How old were you when you first had sex</td>
</tr>
<tr>
<td>How long were you sexually involved with this partner?</td>
</tr>
<tr>
<td>When was the last time you had sex with partner?</td>
</tr>
<tr>
<td>How often did you use condoms with this partner</td>
</tr>
</tbody>
</table>
Exploring DDI RepresentedVariable

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

**Zimbabwe Study**
- Month and year of first sex with this partner

**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)**

**Date of first sex with partner**

**RepresentedVariable**
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!!!