BigSurv18
BIG DATA MEETS SURVEY SCIENCE

ESRA

OCTOBER 25–27, 2018 ■ BARCELONA, SPAIN

A Joint Modelling Approach in SAS to Assess Association between Adult and Child HIV infections in Kenya

Elvis Karanja Muchene | University of Nairobi, Kenya | Mobile: +254713900005 | Email: elviskaranja@gmail.com

Abstract

Recent studies have adopted a joint modelling approach as a more stout technique in studying outcomes of interest simultaneously especially when the interest is in the association between two dependent variables. This has been necessitated by the fact that modelling such outcomes separately often leads to biased inferences due to existing possible correlations especially in medical studies. This paper demonstrates the application of linear mixed modelling approach using SAS analysis software to evaluate the correlation between adult and child HIV infections for each county in Kenya, while adjusting for several predictors of interest. Using HIV data extracted from the Kenya open data website for the year 2014, we visualize on each county the HIV prevalence on the Kenyan map. High infection incidences are observed for counties located in Nyanza province. We further fit a joint model for the two outcomes of interest using the linear mixed models approach to capture possible correlation between the two outcomes for each county. Results indicate that there is a correlation between infections in adults and children. Further, there is a significant effect of ART coverage, adults and children in need of ART and number of people undergoing testing voluntarily. Researchers or students who have little understanding in application of linear mixed models, both theoretical understanding and practical analysis in SAS as well as application on real datasets, will find this article useful. Findings from this article would interest the health sector, practitioners and other institutions working in HIV related interventions

Keywords: Antiretroviral Therapy (ART), HIV, joint modelling, linear mixed model, Repeated measures, SAS.

Introduction

- The most serious HIV and AIDS epidemic in the world is in sub-Saharan Africa
- However, It is possible to end this epidemic in each region and entire population (UNAIDS, 2014)
- Between 2005-2013, Sub Saharan Africa accounted for 74% of all the people dying from AIDS related causes.
- However, AIDS related deaths fell by 39% in Sub Saharan Africa during same duration.

Problem Statement

- HIV AIDS related interventions continue to be put into place by NGOs and other institutions in a bid to reduce the prevalence in sub Saharan Africa.
- It is important to ensure that the design and implementation of such interventions are informed by facts for the envisioned goal to be achieved.
- HIV AIDS prevalence and concomitant significant contributing factors are some of the insights that remain vital.
- This calls for robust analysis of existing data and information to form the basis of decision making.
- In such studies and medical studies in general, its impossible to ignore possible correlations which would lead to biased inference.
- The study therefore seeks to analyse HIV data from Kenya taking into account such correlations and predictors of interest.

Data Description

- Kenyan government, in its commitment to proactively inform its citizens and be accountable, makes public government datasets accessible for free to the public in easy and reusable formats, through the Kenya Open Data Portal.
- We extract 2014 Kenya's HIV AIDS profile data by county from this portal, accessible through http://www.opendata.go.ke/datasets/county-hiv-profiles-2014
- Covariates of interest are;
  - Number of adults and children in need of anti-retroviral therapy (ART),
  - Number of adults and children receiving ART treatment,
  - ART coverage
  - Gender specific HIV prevalence
  - Number undergoing voluntary testing

Figure 1: HIV Prevalence

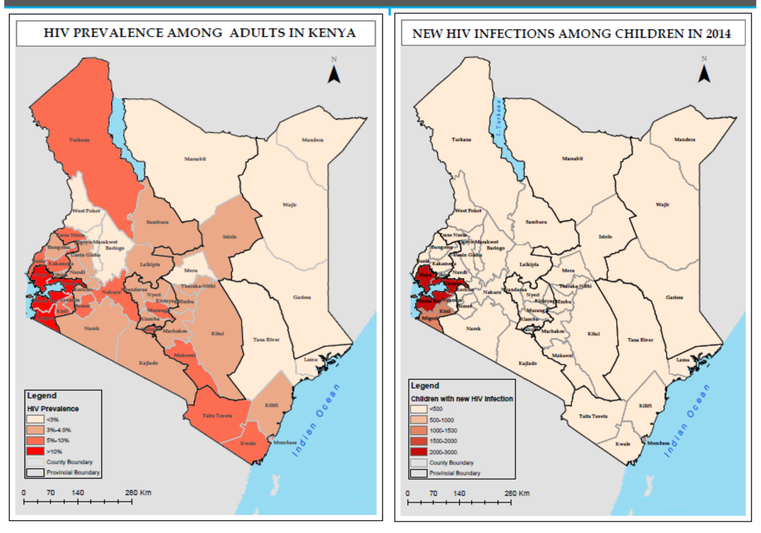
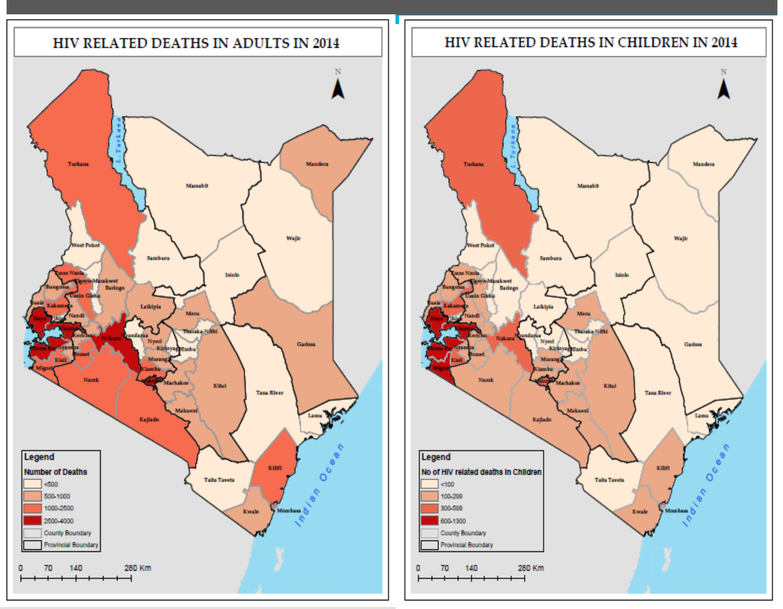


Figure 2: HIV Related Deaths



Joint Modelling-Linear Mixed Model

- Basic assumption is that the data are linearly related to unobserved multivariate normal random variables.
- Standard linear model takes the form; $y = X\beta + \varepsilon$
- The mixed model extends the standard linear model by allowing for a more flexible specification of the covariance matrix of ε.
- Allows for one to capture both the correlation and heterogeneous variances under the assumption of normality
- Comes from the fact that it contains both the random effects parameters, γ and fixed effects parameters β, and takes the form;

$$y=X\beta+Z\gamma+\varepsilon$$

- y denotes vector of observed outcomes of interest (i.e. Child and adult infection rates for this study)
- β denotes fixed effects with design matrix X, γ random effects with design matrix Z
- Under the assumption that γ and ε are normally distributed, the variance of *y* is of the form

$$V = ZGZ' + R$$

- *G* & *R* are covariance structures to be specified

SAS Mixed Procedure-PROC Mixed

- PROC Mixed fits wider class of mixed models, and thus a further generalization of GLM.
- Some basic statements used in our analysis include
  - **CLASS**-names variable classification for analysis
  - **MODEL**-names single dependent variable & fixed effects,
  - **RANDOM**-to incorporate random effects
  - **TYPE**-specifies covariance structure of G
  - **SUBJECT**-identifies the subjects in the model
  - **NOBOUND**-requests removal of boundary constraints on covariance parameters
  - **NOINT**-requests no intercept is included in the model
  - **DDFM**-specifies method of computing denominators degree of freedom for the tests of fixed effects resulting from MODEL statement
  - **GCORR**-displays correlation matrix that corresponds to estimated G matrix

Final SAS Statement

```
PROC IMPORT OUT= WORK.hiv
            DATAFILE= "D:\Documents\To DO\HiV
Data\Elvis\Newlong_hiv.csv"
            DBMS=CSV REPLACE;
RUN;
PROC SORT DATA=WORK.hiv;
BY COUNTY_NAME;
RUN;

data hiv;set hiv;
logresp=log(value);
logadult=log(adults_in_need_of_art);
logart=log(art_coverage);
logartcoverage=log(art_coverage);
logchildrenart=log(children_in_need_of_art);
logmen=log(men_prevalence);
logwomen=log(women_prevalence+1);
logvct=log(volunteer_testing);
run;run;

ods pdf file='D:\Documents\To DO\HiV
Data\Elvis\Elvis_SAS.pdf';

proc mixed data=work.hiv nobound  ;
class county_name   variable;
model logresp = logadult*variable logart*variable
logchildrenart*variable logmen*variable
logwomen*variable logvct*variable/s noint ddfm=residual;
random variable/ subject=county_name type=csh g gcorr  ;
run;
```

Observations during Model fitting

- Dependent on the statement specified in SAS, Convergence criteria may fail to be met.
- Main reason was due to the Hessian matrix, which would fail to be positive definite.
- Hessian matrix is a second order partial derivative of a scalar function.
- Its used to compute standard errors of covariance parameters.
- We resolved this issue by;
  - Rescaling of predictor variables (Log scale)-variable transformation
  - Respecified the random part of the model by specifying covariance structure as Compound symmetry (CSH) instead of Unstructured
- There was also a possibility to run a population averaged model instead of mixed models, had the initial approaches failed
- The best fit model had the smallest AIC value.

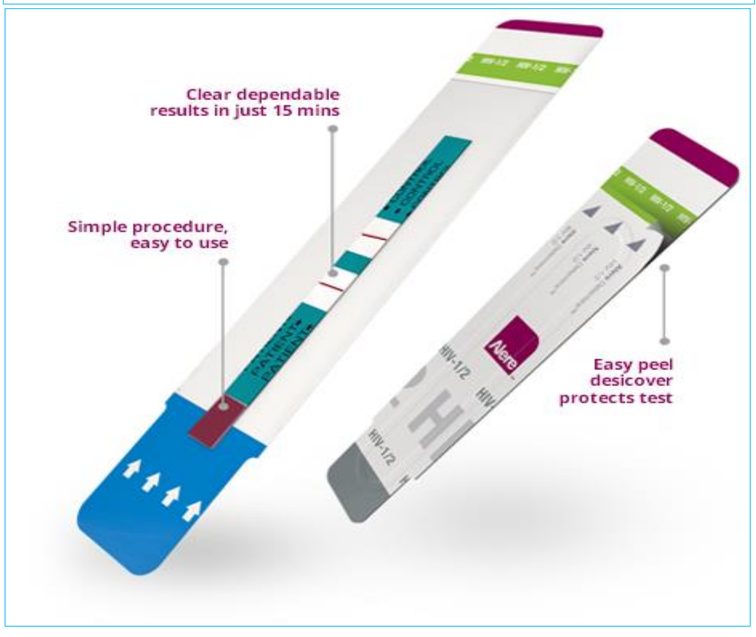


Table 1: Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F-Value	Pr>F
Adults in need of ART	2	80	4.53	0.0137
Children in need of ART	2	80	2.18	0.1192
ART coverage	2	80	7.24	0.0013
Male prevalence	2	80	2.23	0.1142
Women prevalence	2	80	4.06	0.0210
Voluntary testing	2	80	6.67	0.0021

Conclusion.

- Correlation between Adult and Child HIV infections was almost equal to 1.
- This is an indication that there exists a strong positive correlation between adult HIV infections and Child HiV infections.
- In other words, if an adult has a HIV infection, its highly likely that the child has HIV infection too.
- From the type 3 fixed effects, Adults in need of ART, ART coverage, Women prevalence and Voluntary testing are significant predictors of both child and adult HIV

Further Reading

- Kathleen *et al* (2012). Tips and Strategies for Mixed Modeling with SAS/STAT Procedures. SAS Global Forum 2012 (pp. 1-18). NC: SAS Institute Inc.
- SAS Manual
- UNAIDS. (2014). THE GAP REPORT.