Identifying Trends and Patterns in Incidence of AIDS in Bangkok Using Generalised Linear Mixed Models

Pianpool Kirdwichai

Abstract—**Reliable estimation and prediction of AIDS incidence in vulnerable subgroups of the population is recognised as important for efficient planning and allocation of resources and to improve the success of early intervention programmes. This paper addresses this reliability issue and explores the use of modern regression methods in modelling annual AIDS incidence while accounting for non-normality, nonlinearity and correlated responses. In particular, linear and non-linear marginal regression models are considered when the normality assumption is valid. Otherwise generalised linear mixed regression methods are used to develop models for estimation and prediction of incidence for various subgroups of AIDS patients. Annual incidence data from Group Plans and Information AIDS TB and STIs Control Division of Bangkok Health Center and covering the period 2005-2015 is used to build the models and to quantify and compare trends and patterns in incidence for different subgroups of AIDS patients in Bangkok.**

Although the incidence of AIDS in Thailand has been declining over the past decade, the rate of decline is not the same for certain subgroups of the population. The findings show that the annual relative incidence rate is around 80% overall in Bangkok but that the decrease in incidence is slowing down in young men, older people and male prisoners. This latter group have been identified as a key group left behind in prevention strategies. On the other hand the rate of decline appears to be faster in people who inject drugs than the rest of the population.

Index Terms—**AIDS, conditional model, generalised linear mixed regression, non-linear mixed regression, marginal model.**

I. Introduction

THE Pacific region has the second largest number of individuals in the world living with HIV of which approximately 9%, or roughly $445,000$, live in Thailand. Although early intervention programmes in Thailand has led to considerable success in prevention and treatment of AIDS in recent years, with a 34% decline in AIDS related deaths in 2014 relative to 2005 and a 5% decline in new infections between 2010 and 2015, yet $20,492$ people died of AIDS-related illnesses in 2014 [1], [2]. Globally, key subgroups of the population that continue to be left behind in prevention strategies and have much higher risk of HIV infection include young women aged $15-24$ years, sex workers, people who inject drugs, transgender people, prisoners and gay men and other men who have sex with men. Yet few countries have consistently implemented prevention strategies specifically tailored to these subgroups [1].

Manuscript received March 9, 2017; revised March 28, 2017. This research was funded by King Mongkut's University of Technology North Bangkok. Contract no. KMUTNB-60-GEN-032.

The author is Assistant Professor with the Department of Applied Statistics, Faculty of Applied Science, KMUTNB, Thailand e-mail: pianpool.k@sci.kmutnb.ac.th.

The rate of decline in HIV infections in Thailand is thought to be slowing down in key affected subgroups who are more at risk and more vulnerable than others, such as adolescent girls, sex workers and older people [2]. Unprotected sex still accounts for the majority of new adult HIV infections in Thailand. Of all new infections, almost half were among men who have sex with men (MSM), male sex workers (MSW), and transgender people (TG), making these subgroups a priority for prevention work nationally. Within Bangkok, HIV prevalence in MSM is approximately three times the national rate of 7% for this population. Prevention programmes in Bangkok have also failed to reach as many young MSM (compared to older MSM) resulting in younger MSM being less likely to know where to obtain an HIV a test, or understand their risk [3]. While another large at-risk population in Bangkok that are being left behind in prevention and treatment is transgender men and women (TG), nevertheless the group with the highest AIDS prevalence is people who inject drugs (PWID) [4]. HIV can spread rapidly among people who inject drugs and numbers continue to increase largely because harm reduction services for PWID, such as needle and syringe programmes, still are not adequately available [5]. Another priority subgroup are female sex workers (FSW) [4]. Although AIDS prevalence in FSW is far less than in MSW, 12% of incident cases nationally were FSW and their clients [5] and that this figure is exceptionally higher (20%) in Bangkok [6].

In this paper, annual incidence data from Group Plans and Information AIDS TB and STIs Control Division of Bangkok Health Center, covering the period 2005-2015 [7], is used to quantify and compare trends and patterns in AIDS incidence in Bangkok, with particular focus on key affected population subgroups. Profile plots of the Bangkok AIDS incidence data in Figure 1 illustrate highest incidence for men aged between 25 and 44 and a fairly consistent decrease in incidence across all age groups and both genders (see also Figure 2). Figures 3-5 illustrate profile plots of Bangkok AIDS incidence for selected occupation groups and cause of infection. Figures 3 and 4 show that general workers, unemployed males and female domestics have the highest incidences, and in addition, supports the Thai National Aids Committee Report [2] of a decrease in the decline rate in incidence for prisoners. On the other hand, the incidence rates for both male and female sex workers are seen to be comparatively low. With regards to cause, see Figure 5, the highest incidence by far is for the group whose infections were as a result of sex, and there are indications in the data supporting the conclusion [4] that HIV infection among MSM is reaching high levels without any indication of a

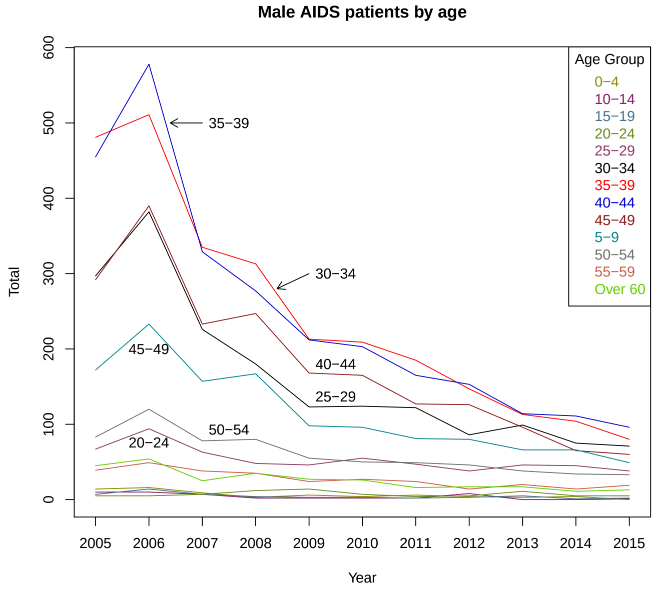

Fig. 1. Profile plot of annual AIDS incidence (2005-2015) for males in Bangkok by age.

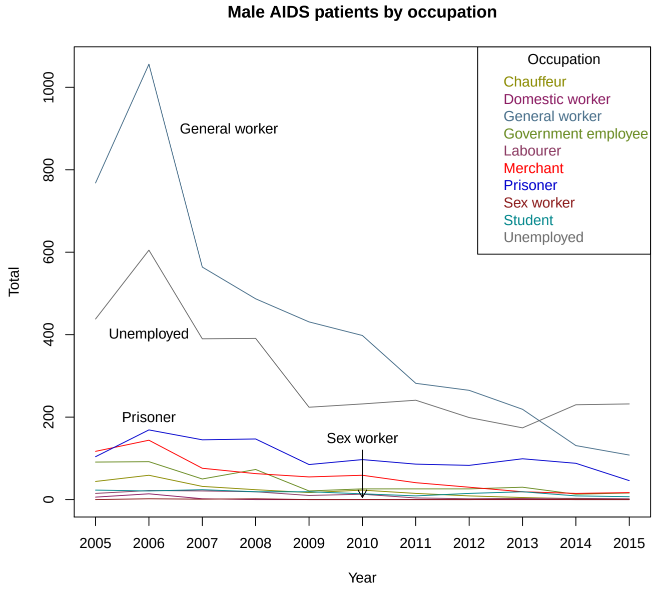

Fig. 3. Profile plot of annual AIDS incidence (2005-2015) for males in Bangkok by occupation.

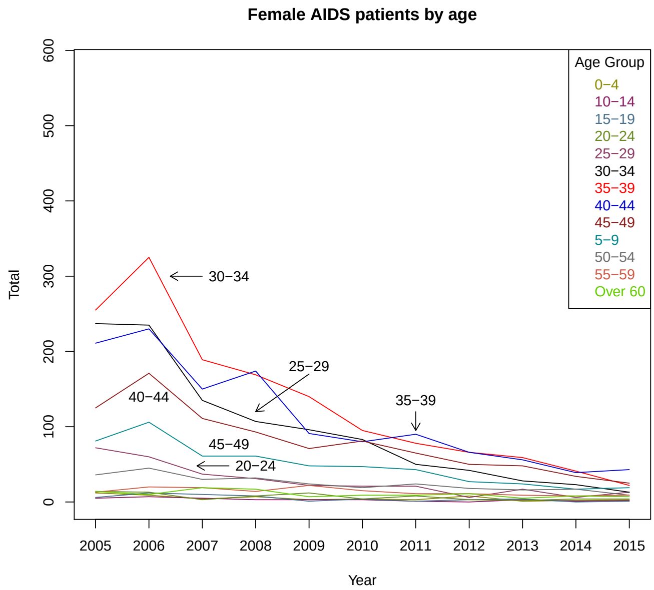

Fig. 2. Profile plot of annual AIDS incidence (2005-2015) for females in Bangkok by age.

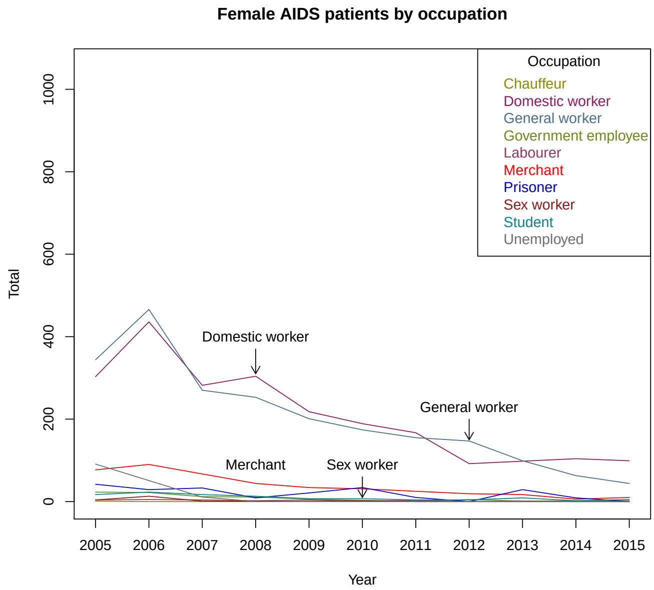

Fig. 4. Profile plot of annual AIDS incidence (2005-2015) for females in Bangkok by occupation.

decline, especially for those MSM in large cities or popular international tourist locations. These patterns will be further explored and quantified in this paper.

UNAIDS plans to reduce new infections globally to $500,000$ by the year 2020 is underpinned by evidence-based mechanisms to inform their prevention programmes [1]. Thailand has also adopted evidence-based decision making in their efforts to prioritize areas for intervention and to allocate resources effectively and efficiently [4]. This paper seeks to further promote and reinforce the evidence-based approach by proposing and illustrating the use of efficient modelling techniques in better understanding the incidence rates of AIDS, in reliably quantifying these rates and in efficiently predicting incidence for subgroups of the population.

II. Method

Longitudinal data, such as the Bangkok AIDS incidence, are typically serially correlated and require analysis methods that allows for this covariance structure and ensures valid estimation of standard errors. Models for correlated count data have been used [8], for example, to investigate trends in US polio incidence. The two most common regression methods proposed for exploring the relationship between a response variable with a within-group correlation structure and a set of independent variables are the conditional model, which is based on use of random effects, and the covariance pattern or marginal model, which directly models the covariance structure of the responses [9]. The conditional model assumes the relationship between the response and

ISBN: 978-988-14048-3-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

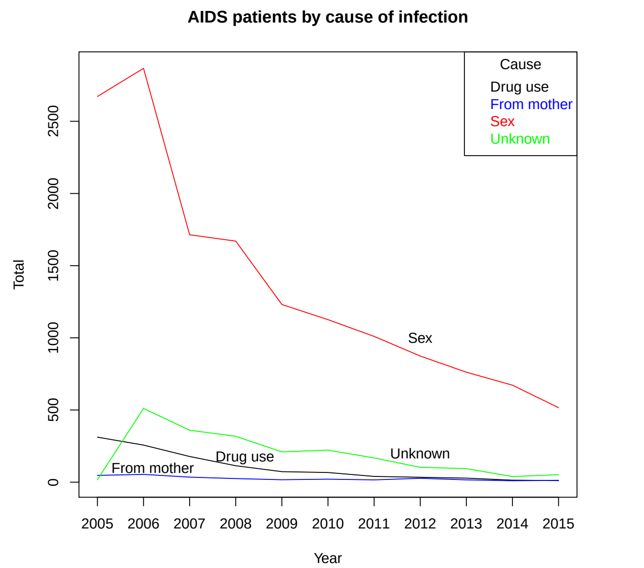

Fig. 5. Profile plot of annual AIDS incidence (2005-2015) in Bangkok by cause of infection.

independent variables is common to all groups (the so-called fixed effects) and accommodates the covariance structure by allowing some model parameters to vary between groups through the use of random effects. As this model incorporates but fixed and random effects in its specification, it is also commonly referred to as a mixed model. On the other hand, the correlation is modelled directly in the marginal model via use of a block diagonal error covariance matrix.

Consider N groups of correlated data, let $y_{i,t}$ denote the t^{th} observation within the i^{th} group, $t = 1, \ldots, n_i$, and suppose the covariance matrix of the vector of responses in the i^{th} group $\boldsymbol{y}_i = (y_{i,1}, \ldots, y_{i,n_i})^T$ is denoted by $\sigma^2\boldsymbol{V}$, for $i = 1, \ \ldots, N$. Then assuming independence between groups, a marginal regression model for the responses is

$$\boldsymbol{y}_i = \boldsymbol{f}(\boldsymbol{x}_i, \boldsymbol{\beta}) + \boldsymbol{\epsilon}_i, \ i = 1, \ \ldots, N \ , \qquad (1)$$

where the vector $\boldsymbol{f}(\boldsymbol{x}_i, \boldsymbol{\beta}) = (f(\boldsymbol{x}_{i,1}, \boldsymbol{\beta}), \ldots, f(\boldsymbol{x}_{i,n_i}, \boldsymbol{\beta}))^T$ for some function f of the independent variables $\boldsymbol{x}$ and the regression parameter vector $\boldsymbol{\beta}$, and $\boldsymbol{\epsilon}_i$ is a vector of correlated errors with covariance matrix $\sigma^2\boldsymbol{V}$. Here f is a nonlinear function in at least some of the parameters in $\boldsymbol{\beta}$ and this is therefore a nonlinear marginal model. The nonlinear marginal model, which can be viewed as a hierarchical model that generalizes both the linear marginal model and the nonlinear regression model for independent data [10], is used in this paper to model the Bangkok AIDS incidence data.

Parameter estimation and inference for the model in equation 1 is based on the normality assumption for the responses. Alternatives considered in this paper when this assumption is invalid belong to the class of generalized linear mixed models (GLMMs). GLMMs are a powerful class of statistical models that can cope with a wide range of response distributions, including binomial, Poisson and negative binomial and, like the marginal model, are useful in a wide-ranging scenarios where observations are not independent but obey some underlying correlation structure, as is the case when data are collected sequentially in time. In addition, GLMMs can potentially allow for complex grouping structures and missing data, and are increasingly being implemented in freeware such as R [11].

A GLMM combines the features of a generalized linear model (GLM) and a mixed model. A GLM consists of a linear predictor $\eta = \boldsymbol{x}\boldsymbol{\beta}$, a monotonic mapping between the mean of the response and the linear predictor and a response in the exponential family of distributions

$$g(y) = exp\left\{\frac{y\theta - b(\theta)}{\phi} + c(y, \phi)\right\} \ .$$

Here b and c are known functions, the parameter θ is called the natural (canonical) parameter and ϕ is a scale parameter. The mean and variance of the response are related to the components of the density g: $E(Y) = \mu = b'(\theta), Var(Y) = \phi b''(\theta)$, where prime denotes derivative. The relationship expressing θ as a function of μ is known as the natural link or the canonical link function. A GLMM is an extension of a GLM in which the linear predictor is given by

$$\eta = \boldsymbol{x}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} \ ,$$

where $\boldsymbol{Z}$ is a design matrix for the random effects $\boldsymbol{u}$, which is often assumed to be normally distributed with mean 0 and variance matrix $\boldsymbol{G}$. Note that this implies that instead of specifying a distribution for the response y, as in the case of a GLM, a distribution for y conditional on u is now specified.

To examine the incidence of AIDS over time and in the different subgroups, nonlinear marginal regression models, as well as GLMMs, with yearly counts of AIDS patients as response variable, and year and subgroup as independent variables, are fitted and evaluated. The models are fitted using the nlme [12] and lme4 [13] libraries in R. Goodness of fit of the GLMM's are evaluated through a crude test of overdispersion comparing the sum of squared Pearson residuals to the residual degrees of freedom.

III. Results

A consistent decrease in the incidence of AIDS is observed for both males and females, see Figure 6. Denote the incidences for the i^{th} gender by the vector y_i, years by the vector t and let s be the indicator for gender (s=1 if male, 0 if female). A model that provides a good fit to the data is the nonlinear marginal model

$$y_i = (\alpha + \beta * s)e^{(\gamma + \delta * s)t} + \epsilon_i, \ i = 1, 2 \ , \qquad (2)$$

where $\alpha, \beta, \gamma, \delta$ are model parameters and the within gender error vector ϵ is assumed to be multivariate normal with an AR(1) covariance structure. It should be noted that the marginal and conditional linear models of the log-transformed incidence both give similar parameter estimates as the non-linear model but have standardized residuals that are slightly left-skewed. The non-linear model also have the additional benefit of providing parameter estimates that are scientifically meaningful within the context of the problem studied. Here the parameters γ and $\gamma + \delta$ represent the annual relative incidence rate for males and females. Estimates of the annual rates $(95\% CI)$ from the fitted model are $0.817(0.794, 0.841)$ for females and $0.831(0.819, 0.842)$ for males, clearly suggesting a decrease in incidence of approximately 20% per year for both groups. A plot of the fitted model is shown in Figure 6.

Table I provides estimates and standard errors of the relative incidence rates for different age groups and by gender,

ISBN: 978-988-14048-3-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

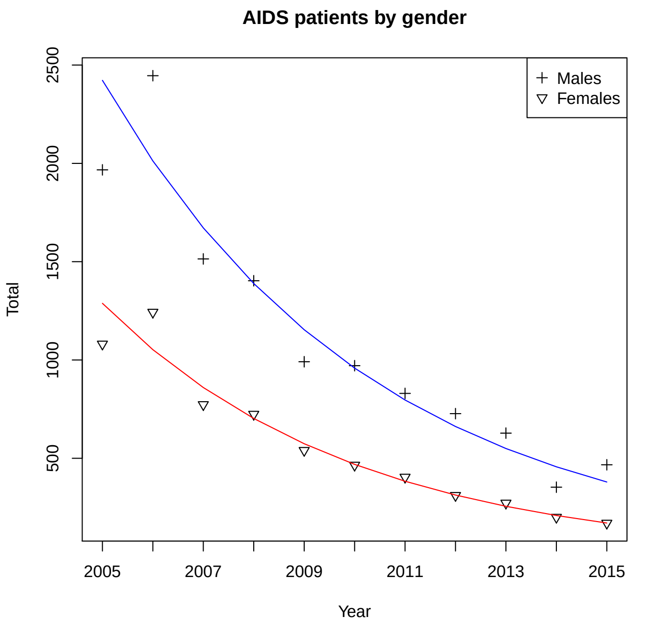

Fig. 6. Scatterplot of annual AIDS incidence in Bangkok. The solid lines are fitted models for the two gender groups.

TABLE I
ESTIMATE (STANDARD ERROR) OF THE ANNUAL RELATIVE INCIDENCE RATE OF AIDS IN BANGKOK BY GENDER FOR DIFFERENT AGE GROUPS.

	Incidence (s.e.)	
Age Group	Male	Female
0-4	0.81 (0.035)	0.86 (0.039)
5-9	0.79 (0.040)	0.84 (0.050)
10-14	0.83 (0.040)	0.80 (0.041)
15-19	0.98 (0.035)	0.87 (0.034)
20-24	0.93 (0.012)	0.80 (0.017)
25-29	0.85 (0.012)	0.76 (0.014)
30-34	0.83 (0.005)	0.80 (0.008)
35-39	0.83 (0.005)	0.83 (0.008)
40-44	0.84 (0.008)	0.85 (0.010)
45-49	0.86 (0.013)	0.84 (0.016)
50-54	0.88 (0.011)	0.89 (0.018)
55-59	0.89 (0.017)	0.92 (0.024)
60+	0.86 (0.017)	0.92 (0.027)

obtained from fitting GLMM's with yearly AIDS counts as response, a random intercept varying at the level of gender, and gender, year, and gender by year interactions as fixed effects. Generalised linear mixed Poisson regression models are fitted initially and tests for overdispersion based on the ratio of Pearson residuals to the model residual degrees of freedom are conducted. For models with evidence of overdispersion, generalised linear mixed negative-binomial models are instead considered. The analysis in the table shows that most of the age groups have relative incidence rates that are consistent with the overall pattern of decline observed. The exceptions are young women aged $25-29$ with a faster rate of decline, and young men aged $15-24$ and older $(50+)$ women and men with rates of decline that are slower than the overall population rate. Indeed for male teenagers, $(15-19)$ age group, the annual relative incidence rate (95% confidence interval) is $0.98(0.911, 1.049)$ and provides evidence of no decline in incidence for this age group. However, it should also be noted that only a very small number of AIDS patients are observed in this 15-19 age group.

Similar analyses by occupation show relative incidence rates consistent with the overall population rate for all subgroups excepting male prisoners, identified as a key subgroup left behind in prevention strategies, male students and the unemployed. With relative rates for the three subgroups that are higher than the overall rate, see Table II, there is evidence that the decline in AIDS incidence is slowing down for these sets of individuals. Also provided in Table II for information are estimates for sex workers, another key subgroup.

TABLE II
ESTIMATE (STANDARD ERROR) OF THE ANNUAL RELATIVE INCIDENCE RATE OF AIDS IN BANGKOK BY GENDER FOR SELECTED OCCUPATIONS.

	Incidence (s.e.)	
Occupation	Male	Female
Prisoner	0.92 (0.009)	0.85 (0.020)
Sex Worker	0.79 (0.134)	0.63 (0.090)
Student	0.91 (0.022)	0.83 (0.028)
Unemployed	0.90 (0.005)	0.32 (0.031)

Estimates of the annual relative incidence rates by cause of infection provided in Table III were obtained from fitting a quasi-poisson generalised linear regression model with yearly AIDS counts as response and cause, year, and cause by year interactions as explanatory variables. The quasi-poisson distribution was used to address the significant over-dispersion in the fitted Poisson regression model. Over-dispersion was tested using the method by Cameron and Trivedi [14], implemented in the R library AER [15]. Notable is the relative incidence rate (95% CI) of $0.71(0.628, 0.792)$ for people who inject drugs, suggesting that the rate of decline in incidence is a faster than the rest of the population for this subgroup.

TABLE III
ESTIMATE (STANDARD ERROR) OF THE ANNUAL RELATIVE INCIDENCE RATE OF AIDS IN BANGKOK BY CAUSE OF INFECTION.

Cause	Incidence (s.e.)
Drug Use	0.71 (0.042)
Mother	0.86 (0.109)
Sex	0.84 (0.072)
Unknown	0.86 (0.078)

IV. CONCLUSION

This paper explored the use of generalised linear and nonlinear mixed regression methods that explicitly models the correlation in longitudinal data in quantifying trends in incidence rates. The availability of functions in easily accessible software such as R to fit generalised linear and nonlinear mixed models is leading to increasing and wide implementation of these methods. However there are caveats associated with usage. For instance, while parameters in Poisson regression models are readily interpretable, the inflexibility of Poisson regression in modelling the variance can result in poor inference. On the other hand, the negative binomial model maintains much of the interpretation of the Poisson model and allows a more flexible variance structure but, as confirmed in the work here, is computationally much more difficult to fit in practice.

Although much progress has been made worldwide to reduce the incidence of AIDS, it is estimated that between

ISBN: 978-988-14048-3-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

1.7 and 2.2 million adults aged 15+ were newly infected with HIV in 2015, partially because of inadequate access to prevention services and information for certain subgroups of the population [4]. In response, it has been acknowledged that an evidence-based approach is key. The work in this paper falls within this evidence-based framework. In particular, reliable models of the trend and pattern of incidence may be used to predict numbers of AIDS patients overall, as well as in key subgroups, while reliable incidence estimates will better inform the planning and allocation of resources and provide quality measures of the effectiveness of prevention programs.

REFERENCES

[1] UNAIDS, “Prevention gap report,” http://www.unaids.org/sites/default/files/media_asset/2016-prevention-gap-report_en.pdf, 2016.
[2] Thai National Aids Committee, “Thailand aids response progress report 2015,” http://www.unaids.org/sites/default/files/country/documents/THA_narrative_report_2015.pdf.
[3] UNESCO, “Young people and the law in asia and the pacific: a review of laws and policies affecting young people’s access to sexual and reproductive health and hiv services,” http://unesdoc.unesco.org/images/0022/002247/224782E.pdf, 2013.
[4] Thai National Aids Committee, “National aids strategy for 2012 to 2016,” 2016.
[5] Thai National Aids Committee, “Thailand aids response progress report 2014,” http://www.unaids.org/sites/default/files/country/documents/THA_narrative_report_2014.pdf.
[6] C. Manopaiboon and D. Prybylski and W. Subhachaturas and S. Tanpradech and O. Suksripanich and U. Siangphoe and L. G. Johnston and P. Akarasewi and A. Anand and K. K. Fox and S. J. Whitehead, “Unexpectedly high hiv prevalence among female sex workers in bangkok, thailand in a respondent-driven sampling survey,” *International Journal of STD and AIDS*, vol. 24, pp. 34-38, 2013.
[7] GIDS TB and STIs Control Division, “Summary of AIDS patients in Bangkok 2015,” http://www.bangkok.go.th/aids/index
[8] S. L. Zeger, “A regression model for time series of counts,” *Biometrika*, vol. 75, no. 4, pp. 921-929, 1988.
[9] J. C. Pinheiro, “Conditional versus marginal covariance representation for linear and nonlinear models,” *Austrian Journal of Statistics*, vol. 35, no. 1, pp. 35-44, 2006.
[10] D. M. Bates and D. G. Watts, “Nonlinear regression analysis and its applications, ” *Probability and Statistics*, Wiley, New York, 1988.
[11] R Core Team, “R: A Language and Environment for Statistical Computing,” *R Foundation for Statistical Computing*, Vienna, Austria, http://www.R-project.org/, 2016.
[12] J. Pinheiro and D. Bates and S. DebRoy and D. Sarkar and R Core Team, “nlme: Linear and Nonlinear Mixed Effects Models,” http://CRAN.R-project.org/package=nlme, 2016.
[13] D. Bates and M. Mächler and B. Bolker and S. Walker, “Fitting Linear Mixed-Effects Models Using,” *Journal of Statistical Software*, ” doi= 10.18637/jss.v067.i01, vol. 67, no. 1, pp. 1-48, 2015.
[14] A. Cameron and P. Trivedi, “Regression-based tests for overdispersion in the Poisson model,” *Journal of Econometrics*, vol. 46, no. 3, pp. 347-364, 1990.
[15] C. Kleiber and A. Zeileis, “Applied Econometrics with R,” Springer-Verlag, New York, https://CRAN.R-project.org/package=AER, 2008.

ISBN: 978-988-14048-3-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)