2011
2012

FACULTY OF SCIENCES

Master of Statistics: Biostatistics

Masterproef

Joint modelling of HBV and HCV infections from cross-sectional serological data

Promotor :
Prof. dr. Ziv SHKEDY

Promotor :
Dr. KAATJE BOLLAERTS

James Orwa

Master Thesis nominated to obtain the degree of Master of Statistics , specialization Biostatistics

Maastricht University

universiteit hasselt
KNOWLEDGE IN ACTION

De transnationale Universiteit Limburg is een uniek samenwerkingsverband van twee universiteiten in twee landen: de Universiteit Hasselt en Maastricht University

universiteit hasselt
KNOWLEDGE IN ACTION

Maastricht University

Universiteit Hasselt | Campus Diepenbeek | Agoralaan Gebouw D | BE-3590 Diepenbeek
Universiteit Hasselt | Campus Hasselt | Martelarenlaan 42 | BE-3500 Hasselt

2011
2012

FACULTY OF SCIENCES

Master of Statistics: Biostatistics

Masterproef

Joint modelling of HBV and HCV infections from cross-sectional serological data

Promotor :
Prof. dr. Ziv SHKEDY

Promotor :
Dr. KAATJE BOLLAERTS

James Orwa

Master Thesis nominated to obtain the degree of Master of Statistics , specialization Biostatistics

Maastricht University

universiteit hasselt
KNOWLEDGE IN ACTION

Certification

This is to certify that this report was written by James Orwa under our supervision.

Signature: _____ Date: _____

Prof. dr. Ziv SHKEDY **Internal Supervisor**

Signature: _____ Date: _____

Dr. Kaatje BOLLAERTS **External Supervisor**

Thesis submitted in partial fulfilment of the requirements for the degree of Master of Statistics: Biostatistics.

Signature: _____ Date: _____

James Orwa **Student**

Acknowledgements

Thanks to the Almighty God for his unlimited blessing throughout this masters study period. I am deeply indebted to my supervisors Prof. dr. Shkedy Ziv and Dr. Kaatje Bollaerts for their guidance, suggestions and advice while working on this project. It was a real pleasure to work with them.I would like to express my heartfelt appreciation for my professors at Hasselt University, my family members and my friends. Especially, I would like to thank the woman who shares my life, Rehema. Her patient love enabled me to complete my studies.

Lastly, I cannot forget to thank Institute of Public Health,Brussels(Belgium) for allowing me to use their data and Vlaamse InterUniversitaire Raad (VLIR) for the financial support during my study period.

James Orwa
University of Hasselt
Belgium, September, 2012

List of Abbreviations

IDU - Injecting Drug User

HCC - Hepatocellular carcinoma

EU - European Union

DNA - Deoxyribonucleic acid

GLMM - Generalized Linear Mixed Model

HIV- Human Immuno Deficiency Virus

MSOC/MASS- Medical Social Centres for drug users

AC- Non-Residential day care centre

GGG, WGC- Centres for mental health care

CIC- Crisis Intervention Centre

TG- Therapeutic Community

PAAZ- Psychiatric unit within general hospital

PH- Psychiatric hospital

Abstract

The aim of this project is to study the co-infection by regressing marginal association and subject heterogeneity of hepatitis C virus (HCV) and hepatitis B virus (HBV) on behavioral risk factors among drug users within drug treatment centers and prisons in Belgium, using a joint modeling approach that deals with multivariate nature of the response.

Using marginal(Bivariate Dale Model (BDM),Bivariate Probit Model(BPM) and Alternating Logistic Regression(ALR)) models, the association measures between HCV and HBV infections estimated at individual level (cluster) in terms of odds ratios and correlation coefficients was regressed against behavioral risk factors. Shared random-effects models that take into account the individual heterogeneity in the acquisition of the infections were fitted as well.

The analysis used cross-sectional data from 972 drug users who agreed to participate in a sero-behavioural study between 2004-2005. The results showed that the infections are positively associated within individuals (BDM; OR=1.87(95% C.I:1.39, 2.34), ALR; OR=1.45(95% C.I=0.91, 0.99), BPM; ρ=1.21(95% C.I=0.89,1.52)). The variance of the individual random effects is positive (σ_b^2=2.09 (95% C.I: 1.06,3.59)) indicating that there is significant individual heterogeneity in the acquisition of the infections.

Known risk factors for the co-infection were found to be; gender, educational level, current IDU, ever been to prison, age at test and sharing of sniffing materials.

KEYWORDS: Current status data, HCV and HBV co-infection, odds ratio, individual heterogeneity, marginal models.

Contents

Acknowledgements ii
List of Abbreviations iii
Abstract iv

1 Introduction **1**

2 Data description **3**
2.1 Study population 3
2.2 Serology 4
2.3 Exploratory data analysis 4
2.4 Socio-demographic characteristics 8
2.5 Drug use and risk behaviors 8

3 Methodology **11**
3.1 Statistical analysis 11
3.1.1 Bivariate Dale Model (BDM) 11
3.1.2 Alternating Logistic regression(ALR) 12
3.1.3 Bivariate Probit model (BPM) 13
3.1.4 Shared Random effects model 13

4 Results **15**
4.1 Marginal models 15
4.1.1 Basic model 15
4.1.2 Influence of other risk factors 16
4.2 Shared Random effects model 19

5 Discussion **21**

6 Appendix **27**

List of Figures

1 Histogram and density estimates for age at test, age at first use of drug of illegal and exposure time . . . 7
2 Bar plots with the observed joint probabilities for HBV and HCV infections . . 8
3 Overall prevalence of HCV and HBV infections . . . 10
4 Confidence interval plot for BDM and ALR models for the models depending on the risk factors . . . 18
5 Confidence interval plot for BDM models for the constants odds ratio model . . 19
6 Confidence interval plot for the variance . . . 20

List of Tables

1 Demographic and behavioural of the population . . . 6
2 Demographic and behavioural of the population (continued) . . . 7
3 Univariate association of the risk factors for HBV and HCV infections . . . 9
4 Constant measure of association for different models using different link functions . . . 15
5 AIC values for the BDM and BPM models . . . 17
6 Parameter estimates and standard errors for the association for the best fitting BDM models . . . 18
7 Parameter estimates for the association measures across the levels of the behavioral risk factors . . . 18
8 Parameter estimates for the shared random effect model . . . 19
F.1 Demographic and behavioral characteristics for HBV and HBV status according to risk factors . . . 28
F.2 Parameters estimates and C.I for the Logit links models for HBV infection . . . 29
F.3 Parameters estimates and C.I for the additive and interaction models for HCV infection-logit link . . . 29
F.4 Variables description . . . 33

1 Introduction

Hepatitis B virus (HBV) and Hepatitis C virus (HCV) are major public health concerns worldwide. This is because of shared routes of transmission that is, both are blood borne RNA viruses that replicate rapidly. The Hepatitis C is an infectious disease caused by hepatitis C virus (HCV),which mainly affects the liver. The HCV infection is transmitted through already infected blood contact, particularly associated with the use of syringes , medical poorly sterilized and blood transfusions. It is estimated worldwide that people affected by the hepatitis C range from 130 million to 170 million[1]. The existence of hepatitis C (originally "non-A non-B hepatitis") was postulated in the 1970s and proven in 1989. Hepatitis B is an infectious inflammatory illness of the liver caused by the hepatitis B virus (HBV) that affects hominoidea, including humans. The virus is transmitted by exposure to infectious blood or body fluids such as semen and vaginal fluids, while viral DNA has been detected in the saliva, tears, and urine of chronic carriers. Some of the risk factors for developing HBV infection include getting pricked by infected syringes while working in a healthcare setting, blood transfusions, and dialysis, acupuncture and tattooing. However, Hepatitis B viruses cannot be spread by holding hands, sharing eating utensils or drinking glasses, kissing, hugging, coughing, sneezing, or breastfeeding.

The major route of transmission for both the infection is through contaminated blood contact or blood products and that explains why drug users(DUs) are most at risk either through sharing of injecting materials and or other paraphernalia. In Europe, injecting drug use is a major transmission route for HCV infection. Estimates suggest that around 1 million people who have injected drugs may be living with HCV in the EU today. Typically between 40% and 90% of injecting drug users are infected, and many contract the disease soon after their first injection. This is due to unsafe injecting practices which include sharing of needles/syringes and other injection materials used by already infected peers[5].

The prevalence of HBV and HCV in Europe varies widely among the European countries (EU and EEA/EFTA)[3]. In western Europe the prevalence HCV infection in the general population increases from 0.1% in the North to more than 1% in the South[12]. In Belgium a study in the hospital population showed anti-HCV in 0.87% of the serum samples[13]. For HBV, the prevalence in western Europe was estimated at 5-7% of the general population and 0.5-2% were chronic[14] and in Belgium the prevalence in hospitalized population (including acute infections, recovered and chronic carriers) was estimated at 7.4% [13]. The increase in the prevalence in Belgium has been attributed to the increase use of drug by sharing of injecting materials and other materials like filters, spoons and rinse water have been found also to be a risk factor especially for HCV infection[15] and[16].The prevalence rates of co-infection with HBV and HCV in HIV patients have been variable worldwide depending on the geographic

regions, risk groups and the type of exposure involved which may be different not only from country to country, but also in different regions of the same country. Some studies have also shown that HIV is a risk factor to acquiring either HBV or HCV or both[19].

Since HCV and HBV infections share the same route of transmission, co-infection is likely to be common thing among the susceptible population. The primary concern with HBV/HCV co-infection is that it can lead to more severe liver disease and an increased risk for progression to liver cancer (HCC) which eventually leads to liver related deaths[18]. This study is aimed to model the HCV and HBV co-infection and their dependent on the risk factors among drugs users using marginal and random effects model based on data from drug treatment facilities and prisons spread across Belgium taking into account the clustered nature of the data.

2 Data description

The data contains information on HCV, HBV and HIV among drug users in contact with drug treatment centers and those in prisons interviewed between 1/9/2004 and 30/6/2005 and screened for the same diseases. This is a cross-sectional multi-center sero-behavioral prevalence study in which patients were interviewed within the centers and blood samples collected for screening for the presence/absence of the diseases (HIV,HBC and HVC) at the central laboratory based in the Institute of Public Health in Brussels. Clustered data arise when multiple observations are collected on the same sampling or experimental unit. In this study, the multiple observations arose from a patient blood serum sample that was tested for more than one antigen resulting into a clustered data at a patient level.

2.1 Study population

Patients were recruited into the study if :1) aged between 15-40 years at the time of interview and 2) reported to be using or had used regularly one or more of drugs by any known route of administration (Injecting or Sniffing) in the past 12 months. Those who were eligible and agreed to participate in the study signed two informed consent forms, one to participate in the study and another for the blood to be drawn. A drug treatment center is defined in the Belgium TDI protocol as a recognized center financed by an authority as treatment center (whether or not for its specific assignment towards DUs), that takes care of people with drug problem and provides treatment.

All participants were interviewed by means of a standardized face-to-face interview using questionnaire to obtain information on the patterns of drug use, risk behavior, legal problems, infectious diseases, socio-demographics, use of drug treatment facilities, use of health care services and knowledge and attitude on infectious diseases. A unique identifier to match data from both questionnaire and blood test results was used to keep participants confidentiality and to avoid duplication of data. In total 972 participants had matching identification numbers for both the questionnaire and the laboratory results and were retained for the current analysis.

The following variables were considered to be of interest in this study: Gender, Age in years at the time of interview, Participant's level of Education, Type of center, duration of exposure to drug use , Age at first use of illegal drugs, Sharing of injecting materials , Sharing of sniffing materials, ever been to prison or not and sexual risk behavior (Men having sex with Men(MSM)). Duration of exposure to drug use is defined as the time interval(in years) between the age at first use of drug and the age at test. The study will consider HBV and HCV as a bivariate binary-dependent responses from an individual. See Appendix Table F.4 for the full description of the variables.

Table 1 and 2, show the distribution of the demographic and behavioral characteristics for the population. Note that the samples size for some of the variables may not add up to 972 due to some missing observations in some of the risks factors.

2.2 Serology

For serological data, the current status of the disease depends on the antibody level of the ith subject Z_i, i=1....N and a manufacturers threshold value τ. In that case,the current status of each individual for the jth infection,$j = 1, 2$, is defined by:

$$y_{ij} = \begin{cases} 1 & \text{if } Z_{ij} > \tau_j \quad \textit{seropositive} \\ 0 & \text{if } Z_{ij} \leq \tau_j \quad \textit{seronegative} \end{cases}$$

where τ_j is the infection-specific threshold to classify individuals as either seropositive or seronegative.

The screening for HBsAg, anti-HBC and anti-HBs was perfomed with ETI-MAK-4, ETI-AB-COREK PLUS and ETI-AB-AUK-3 (DiaSorin, Italy) respectively. The subsequent interpretation of the results was done using ETI-MAX 3000 apparatus (DiaSorin, Italy). A test result was classified as HBV infected if either HBsAg and anti-HBC results turns positive or when Anti-HBC and Anti-HBs are positive. Screening for anti-HCV was performed with the $ORTHO^{®}HCV$ 3.0 ELISA. Positive test results were confirmed with a more specific serology while the HIV screening was performed with a combined anti-body/antigen tests. Reactive sera were further tested with an alternative screening assays and further with a confirmatory assay, called InnoLia which can differentiate between HIV-1 and HIV-2 antibodies. A test was classified as reactive if the three test were positive.

2.3 Exploratory data analysis

To understand the nature of the data and variables, descriptive statistics were applied as exploratory tool to describe the socio-demographic characteristics of the participants. A histogram with density plots were applied to explore the distribution in years of drug use career ,age at the time of interview and age at first use of illegal drugs. Chi-square test of independence was applied to identify risk factors associated with each of the infections. The risk factors identified by univariate analysis were fitted into multivariate logistic regression analysis to further describe the risk factors for each infection independently. A bar plot of joint prevalence over the years of exposure was used to explore possible effects of the duration of exposure to drugs

on the infections. Odds ratio and Tetrachoric correlations were used to study the association between HBV and HCV infection .

Table 1: Demographic and behavioural of the population

Factors	**Frequency**	**%**
Gender (N=972)		
Male	777	79.94
Female	195	20.06
Educational level(N=961)		
Lower education	635	66.08
Higher education	326	33.92
Type of Center(N=972)		
AC and MSOC	537	55.25
CGG,TG and WGC	170	17.49
CIC	134	13.79
PAAZ and PH	134	13.48
Sharing injecting materials(N=467)		
Yes	321	68.74
No	146	31.26
Sharing sniffing materials(N=704)		
Yes	520	73.86
No	184	26.14
HIV status(N=972)		
Negative	953	98.05
Positive	19	1.95
HBV Vaccination(N=558)		
Vaccinated	233	41.76
Not vaccinate	325	58.24
Current IDU(N=798)		
Yes	319	39.97
No	479	60.03
Ever IDU(N=821)		
Yes	546	66.50
No	275	33.50
Ever been to prison(N=794)		
Yes	483	60.83
No	311	39.17
Homosexual(MSM)(N=484)		
Yes	11	2.27
No	473	97.73

Table 2: Demographic and behavioural of the population (continued)

Factors	**Frequency**	%
HBV prevalence		
Negative	880	90.53
Positive	92	9.47
HCV prevalence		
Negative	687	70.68
Positive	285	29.32
Median duration of exposure(Q0.25,Q0.75)	23(8-18)	
Median age at test(Q0.25,Q0.75)	29(24-35)	
Median age at first drug use(Q0.25,Q0.75)	15(14-17)	
Tetrachoric correlation for co-infection,ρ	0.55(C.I:0.45-0.66)	
Odds ratio for co-infection,ψ	6.82(C.I:4.26-10.90)	

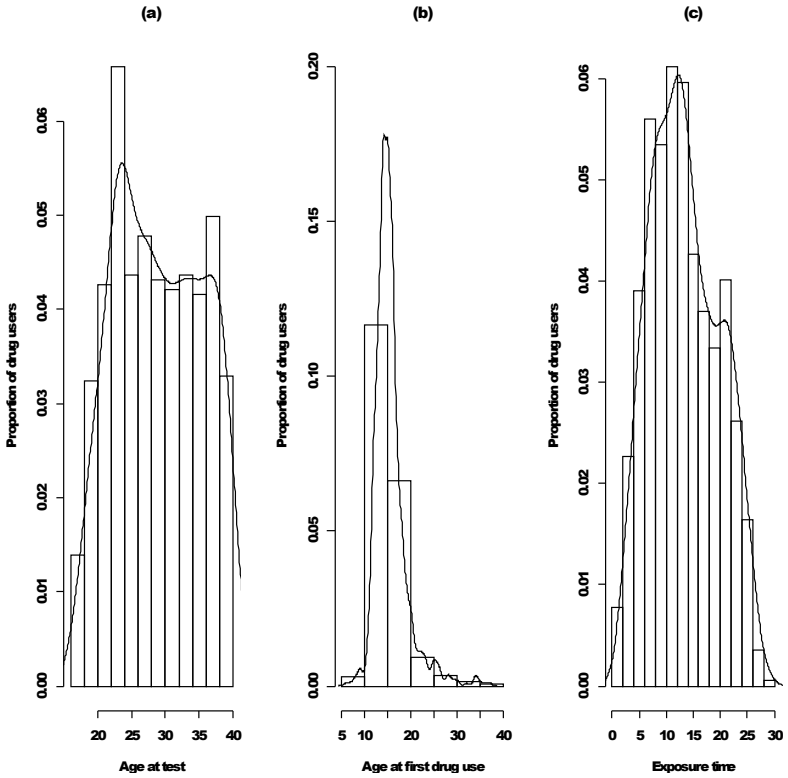


Figure 1: Histogram and density estimates for age at test, age at first use of drug of illegal and exposure time

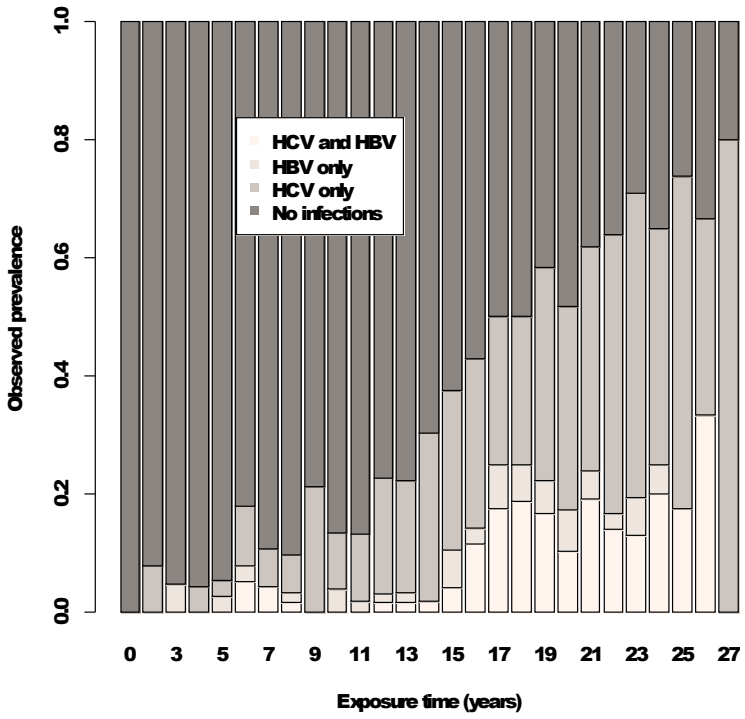


Figure 2: Bar plots with the observed joint probabilities for HBV and HCV infections

2.4 Socio-demographic characteristics

Table 2, shows that, the median age(years) at test was estimated to be 29 (IQR:24-35) and the median age(years) at first use of drugs was 15(IQR:14-17). The histogram and density showing the distribution of age at test and age at first injection are shown in Figure 1. Figure 1(b) shows that majority of this population start injecting drugs early in the course of their life i.e between ages 10 and 15 years; 66.08% of the participants had a lower educational level and this was significantly associated with both HBV and HCV with estimated p-values of 0.0027 and <0.001 respectively(Table 3). Males were 777(79.94%) of the total number of patients, chi-square test of independence did not show any significant association between gender and either HCV or HBV infections. Most of the participants(55.25%) were from AC and Medical Social Assistant Centers for drug use (MSOC) while the rest of the centers had similar number of participants. The type of center had a significant association with both HBV and HCV (p-values of 0.005 and <0.0001 respectively)(Table 3).

2.5 Drug use and risk behaviors

Presented in Table 2 also are the demographic and behavioral characteristics of the participants. It can be shown that, the duration of drug use ranged from 0 to 29 years, the median duration of drug use in years was estimated to be 23(IQR:8-18) years. The bar plot for the prevalence of no

infection, HBV, HCV and co-infection showed that the prevalence of the co-infection increases with the increase in duration of exposure reaching the peak at time 26 years(Figure 1) which suggests possible association of the exposure time and the infections. Univariate analysis(Table 3) showed that educational level,Type of center, vaccination against HBV, being a current injector, ever being an IDU, duration of exposure and age at test were significantly associated with HBV while for HCV all the risk factors apart from gender and MSM were significantly associated with the infection. See Appendix Table F.3 for the description of the multivariate logistic regression analysis.

Table 3: Univariate association of the risk factors for HBV and HCV infections

	HBV		HCV	
	χ^2	p-value	χ^2	p-value
Gender	0.89	0.34	1.05	0.31
Educational level	8.97	0.003	26.90	<0.0001
Center type	17.87	0.0005	39.47	<0.0001
Sharing Injecting material	2.096	0.15	24.17	<0.0001
Sharing sniffing materials	3.77	0.05	9.79	0.0018
HIV status	0.90	0.34	23.03	<0.0001
Vaccination against HBV	10.01	0.0016	14.45	0.0002
Current IDU	9.30	0.0023	101.58	<0.0001
Ever IDU	21.67	<0.0001	166.76	<0.0001
Ever been to prison	27.92	<0.0001	51.97	<0.0001
Homosexual(MSM)	1.13	0.26*	2.12	0.15*
Duration of exposure		<0.0001		<0.0001
Age at test		<0.0001		<0.0001

* Fisher exact test

The tetrachoric correlation was estimated to be 0.55, while the estimated odds ratio between the two infections was 6.82. All the estimates were significant suggesting significant association between HBV and HBC infections(Table 2). Consequently, Figure 3, shows a scatter plot for the prevalence of HBV and HCV infections grouped by the duration of drug use, there is evidence of a positive linear relationship between the prevalence of HBV and the prevalence of HCV infections. These evidence of correlation has to be taken into consideration in order to obtain valid parameter estimates and inference. The joint modeling was therefore used to fit a single model to both response variables (HBV and HCV) simultaneously while taking the correlation between the two into account.

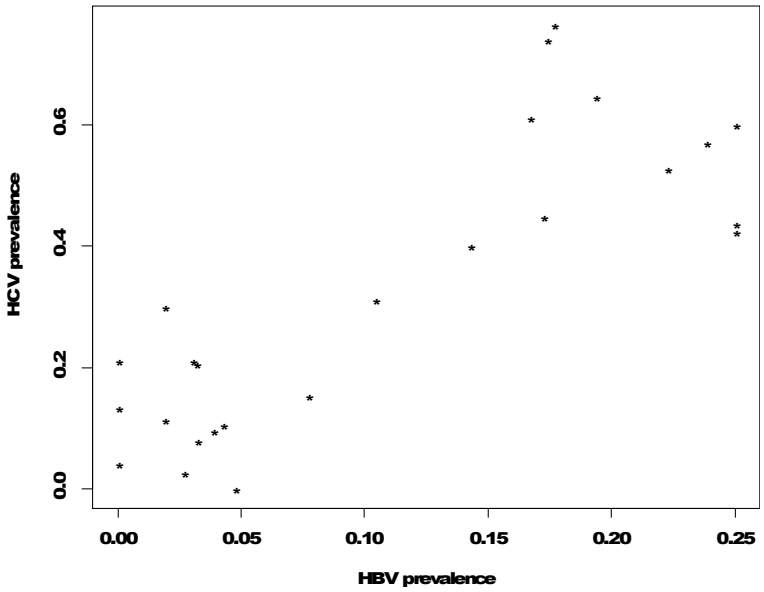


Figure 3: Overall prevalence of HCV and HBV infections

3 Methodology

3.1 Statistical analysis

For each subject i define the vectors representing the serological status for HBV and HCV as $\mathbf{y}_i = (y_{iHBV}, y_{iHCV})$ and the joint probability of co-infection denoted as $P(y_{iHBV} = 1, y_{iHCV} = 1)$. Several methods for analyzing repeated responses focusing on the association between bivariate responses and covariates have been suggested in the literatures. One such approach is the use of marginal models(BDM,BPM and ALR) which allow for inferences about parameters averaged over the whole population ([22], [24] and [9]). Another approach is making use of random effects modeling, which provides inference about variability between subjects ([23] and [10]). The scientific focus in marginal modeling is devoted to two regression models: the first is useful for analyzing marginal means and the second takes into account the correlation between responses of the same individual. This model considers the effect of some explanatory variables on the responses and marginal pairwise local odds ratio [25] or conditional odds ratio [26] for the correlation structure which depends on covariates. Under marginal models, Bivariate Probit and Bivariate Dale models will be the focus while under random effects model, the focus will be on shared random effects model. Specifications of these methods are discussed in detail in the below sections.

3.1.1 Bivariate Dale Model (BDM)

Given the vector of indicators for the dependent variables $\mathbf{y}_i = (y_{iHBV}, y_{iHCV})$ and vector of explanatory variables $\mathbf{x}$, the focus is to model the response as a function of the explanatory variables to determine the dependency of the response on the covariates and also to quantify the degree of association between the pairs of response and their dependency on the covariates [6]. Four potential outcomes can be derived from the set of dependent variable which can be modelled jointly with three systematic components:the marginal $Pr(y_{iHBV} = 1)$ and $Pr(y_{iHCV} = 1)$ and the odds ratio Ψ that describes the dependency between the marginals.

Dale (1986) proposed a family of bivariate response models by decomposing the joint probability into main effects and interactions. The main effects are described by the marginal probabilities of HCV and HBV infections while the interaction is described by the log cross-ratio. The three models in the case of categorical risk factor can be presented as:

$$h(Pr(y_{iHBV} = 1)|x) = \beta_{01} + \beta_{11} log(d_i) + \gamma_{HBV} x_i$$

$$h(Pr(y_{iHCV}=1|x)=\beta_{02}+\beta_{12}log(d_i)+\gamma_{HCV}x_i$$

$$log(OR|x)=\alpha_0+\alpha_1 x_i$$

where log d_i is the logarithm of duration of exposure d_i,α_0 is a constant odds ratio,α_1 odds ratio for the risk factor, $h(.)$ is the link functions, $Pr(y_{iHBV}=1)$ and $Pr(y_{iHCV}=1)$ are the marginal probabilities , x_i is a covariate associated with the HBV and HCV and need not to be the same for the two responses,and γ_{HBV} and γ_{HBV} are the coefficient of the risk factors associated with HBV and HCV infections.

In case x_i is a continuous variable such as age at test, the models can be expressed as:

$$g(Pr(y_{iHBV}=1)|x)=h_1(x)$$

$$g(Pr(y_{iHCV}=1)|x)=h_1(x)$$

$$log(OR|x)=h_3(x)$$

where h_i,$i=1,2,3$ are smooth differentiable functions and $g(.)$ is a link function.
The log odds ratio which describes the dependency between both infections is define as;

$$OR=\frac{\Pi_{11}\Pi_{00}}{\Pi_{10}\Pi_{01}}=\frac{(pr(y_{iHBV}=1,y_{iHCV}=1)pr(y_{iHBV}=0,y_{iHCV}=0))}{(pr(y_{iHBV}=1,y_{iHCV}=0)pr(y_{iHBV}=0,y_{iHCV}=1))}$$

if $OR=1$ indicates both infectious diseases processes behave independently whereas $OR\neq 1$ indicates association between both diseases. For the current analysis, the OR was considered as a constant(Model 1), considered as depending on behavioral risk factor (Model 2) and considered as depending on the log duration and other risk factor (Model 3). Three Link functions; logit, probit and complementary log log were considered and the best model selected based on the AIC values.

3.1.2 Alternating Logistic regression(ALR)

Another method to capture association between categorical responses in terms of odds ratio is a method of Alternating Logistic Regression (ALR)[20]. There are numerous choices for modeling the log odds ratio[21]. For the current analysis,the log odds ratio was specified to be constant within different levels of each level of the behavioral risk factors of HCV and HBV infections. It is worth noting that ALR, likewise BDM, models the association in terms of the odds ratio.

3.1.3 Bivariate Probit model (BPM)

In the bivariate probit model, two separate probit models with correlated disturbances are modeled simultaneously. The basic assumption of this type of model is that current status of the infections are related to the unobserved variables $y_i^* = (y_{i1}^*, y_{i2}^*)$ representing the unobserved variables for HBV and HCV antibody levels. The unobserved variables are assumed to have a bivariate normal density with mean vector $\mu = (\mu_1, \mu_2)^{'}$ and a correlation ρ. The variance-covariance matrix is a 2x2 matrix given by:

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_{y1}^2 & \sigma_{y1}\sigma_{y2} \\ \sigma_{y1}\sigma_{y2} & \sigma_{y2}^2 \end{pmatrix}$$

Further,assume that each subject has p-dimensional vector of explanatory variables $x = (x_0, x_1, ..., x_{p-1})$ with $x_0 \equiv 1$, then the mean is related to the covariates as $\mu_j = \beta_j^{'} x$. In the binary data situation this can be presented with a set of three equations as:

$$\Phi^{-1}(P(y_{iHBV} = 1)) = \beta_{01} + \beta_{11} log(d_i) + \gamma_{HBV} x_i$$

$$\Phi^{-1}(P(y_{iHCV} = 1)) = \beta_{02} + \beta_{12} log(d_i) + \gamma_{HCV} x_i$$

$$log \frac{1+\rho_i}{1-\rho_i} = \alpha x_i$$

Where $\log(d_i)$ is the logarithm of duration of exposure d_i. The Φ is the bivariate standard normal cumulative distribution function. The bivariate probit model models the association in terms of correlation coefficient and uses 'rhobit' link for the association. The ρ_i was considered as a constant (Model 1), depending on other behavioral risk factors(Model 2) or depending on the log duration and other risk factor(Model 3).

3.1.4 Shared Random effects model

Random effects models assume that individuals are different in the way they acquire infections. This may be due to difference in immunity and risk behavior. Some individuals are more susceptible and acquire infection earlier than others. This heterogeneity in acquiring infection can be captured using random-effects. The random-effects approach is a statistical modeling concept which aims to account for heterogeneity, caused by unmeasured covariates such as social environmental factors influences particular to a given subject. A natural way to model dependency within a cluster is through the introduction of cluster-specific random effect. The generalized linear mixed model(GLMM) is the most frequently used random-effects model in the context of discrete repeated measurements [9]. The general GLMM model can be expressed as $Y_i|b_i \sim F_i(\theta, b_i)$ i.e.,conditional on b_i,$Y_i = (y_{iHBV}, y_{iHCV})$ follows a pre-specified distribution F_i depending on the covariates and parameterized through a vector θ of unknown parameters,

common to all subjects, and a vector b_i which is cluster-specific. Since the aim is to assume a joint model specific to each cluster, conditional model will be considered. b_i is assumed to be multivariate normally distributed,$b_i \sim N(0, \sigma_b^2)$ and accounts for the variability not explained by the covariates in the model. A special case of the above model is a shared-parameter model which assumes the same set of random effects for all outcomes specific. The advantage of such shared-parameter models is the relatively low dimension of the random-effects distribution since the dimension of random effects does not increase with the number of new outcomes added[11]. Conditional on b_i the prevalence of HCV and HBV infections assuming shared random intercept model can be modeled as:

$$g(P(y_{iHBV} = 1|b_i)) = \beta_{01} + b_i + \beta_{11} log(d_i) + \gamma_{HBV} x_i$$

$$g(P(y_{iHCV} = 1|b_i)) = \beta_{02} + b_i + \beta_{12} log(d_i) + \gamma_{HCV} x_i$$

b_i is a cluster-specific random intercept that shows how each subject deviates from the overall intercept. Rejection of null hypothesis ($H_0 : \sigma^2 = 0$) implies that infections are correlated at patient level.

In this study we follow the modelling approach of Diamond and McDonald [27] and Keiding et al. (1996) [28] who proposed generalized linear model to fit a parametric model for the prevalence with a linear predictor given by :

$$h_j(d_i) = \beta_{0j} + \beta_{1j} log(d_i)$$

In case that behavioral risk factors are included to this model the linear predictor becomes:

$$h_j(d_i) = \beta_{0j} + \beta_{1j} log(d_i) + \gamma x_i$$

where β_{0j} and β_{1j}(j=1,2) are infection-specific intercepts and slopes , respectively,x_i is a covariate representing behavioural risk factor ,γ is its coefficient which is infection-specific and d_i is the duration of drug exposure in years.

4 Results

4.1 Marginal models

4.1.1 Basic model

The first marginal models were fitted for BDM and BPM which are based on fully-likelihood approach for both constant odds ratio and odds ratio depending on the log duration of exposure. Note that these first models were fitted with only the logarithm of the duration injecting career included in the model as a covariate without any other risk factors. In all the cases,the constant odds ratio model had the least AIC values (Table 5) and was selected as the final model for this particular analysis. Presented in 4, are the parameter estimates of the final models together with the AIC values for the different link functions. The BDM models had the AIC values that were very close to each other eventhough, the probit link had the smallest AIC value. For the ALR which is a quasi-likelihood method, cloglog had the least QIC value while for the shared random effect model,cloglog link function had the least AIC value. The BPM model had no other link function for comparison. However, for the ease of comparison and interpretation in terms of logarithm of the odds ratio for binary responses [29], logit link will be considered in this case and in the subsequent analysis instead of the cloglog and probit. The parameter estimates were found to be similar in all the cases and the 95% confidence intervals show that there was a significant association between the probability of acquiring HBV and HCV infections.

Table 4: Constant measure of association for different models using different link functions

Bivariate Dale Model			
Link function	Logit	Cloglog	Probit
Ψ^{a}(C.I)	1.8664(1.3909,2.3419)	1.8661(1.3906,2.3416)	1.8666(1.3910,2.3422)
AIC	446.49	446.52	446.29
Bivariate Probit Model			
ρ^{b}(C.I)			1.2075 (0.8948,1.5202)
AIC			445.76
Alternating Logistic Rergression			
Ψ^{a}(C.I)	1.4485(0.9118,0.9852)	1.4466(0.9086,1.9847)	1.4751(0.9454,2.0047)
QIC	1557.5164	1551.8203	1571.1517
GLMM-Shared random effect model			
σ_b^{2c}(C.I)	2.0943(1.0582,3.5913)	1.2571(0.6379, 2.1309)	0.7567(0.4147,1.2417)
AIC	1531.25	1526.29	1537.16

[a] Odds ratio
[b] Correlation
[c] Variance

4.1.2 Influence of other risk factors

The above model was extended by adding other risk factors one at a time while adjusting for the log duration of drug use. Three different models categories were fitted for BDM model; model with constants log odds ratio, model with log odds ratio depending on a given risk factor and a model with log odds ratio depending on both log duration of drug use and a given risk factor. The same was also applied to BPM model but in terms of correlation instead of log odds ratio. The AIC values for these models with the best models for each risk factor shown in bold are given in Tables 5 for BDM models and for BPM models where correlation was used instead of log odds ratio. Models with a constant log odds ratio were appropriate for the models containing one of these variables as a covariate in the model; sharing injecting materials ever IDU, homosexual(MSM), ever IDU, HIV status and center. This suggests that there is a significant association of the joint probability of acquiring HBV and HCV infection , however the association does not depend on any of these risk factors and therefore these models could not be used to study the association of the co-infection and the risk factors. The models with log odds ratio depending on the risk factors were appropriate for models containing one of these as a covariate ; educational level, sharing of sniffing materials, current IDU, ever been to prison ,age at the time of interview and gender and was used to study the association of the co-infection and the risk factors. This conclusion was the same for other link functions ; probit and cloglog and also for BPM models.

The final mean structure for the model with association depending on the risk factor includes infection-specific intercepts and slope for log of duration of drug use and one of the behavioral risk factors as additional covariate. The parameter estimates and standard errors for the constant log odds ratio and log odds ratio depending on the duration of exposure for the BDM models are shown in Table 6. It can be shown that there is a significant association in the probability of HBV and HCV co-infections and this association depends on the risk factors apart from age. For the variable age at the time of the interview, there was association between the probability of acquiring HBV and HCV but this does not depend on age. Adjusting for the log duration of drug use, sharing of sniffing materials, is significantly associated with HBV and HCV infection and this co-infection depends on the sharing of the sniffing materials in the last 12 months. This was the case with current IDU and ever been to prison.

Presented in Table 7 , Shows the parameter estimates and the 95% confidence intervals across the levels for each of the behavioral risk factors associated with the co-infection of HBV and HCV for BDM, ALR and BPM models. The BDM model,shows that there is a significant association between HBV and HCV infection for both males and females. The log odds ratio for females is 2.92 with confidence interval of (1.06,4.78) and for males is 1.21 with confidence interval of (0.68,1.75). The associations seems to be stronger for females compared to males.

Consequently, being a current IDU(log(OR)=1.98: C.I=0.94,3.02) shows a stronger association compared to their peers who are not currently injecting drugs(log(OR)=0.92: C.I=0.19,1.66), both levels were significantly associated with the co-infection. Those who have been sharing sniffing materials in the last 12 months had a higher probability of HBV and HCV infections compared to those who have not shared sniffing materials in the last 12 months, however the log odds ratio was significant only for those who have been sharing sniffing materials. The log odds ratio of sharing sniffing materials is 2.69(C.I=1.74,3.69) and for not sharing sniffing materials is 0.82(-0.20,1.85). Ever been to prison was also found to be a significant factor for both infections and both levels showed a significant log odds ratio, however those who had ever been to prison(log(OR=0.89):C.I=0.33,1.45) had a lower odds of getting infected compared to their peers who have never been to prison(log(OR)=2.64: C.I=1.04,4.24). The conclusion was the same for BPM and ALR model results. The confidence interval for the models presented in Table 7 are plotted in figure 4.

Table 5: AIC values for the BDM and BPM models

	BDM			BPM		
Risk factors	Model 1[a]	Model 2[b]	Model 3[c]	Model 1[a]	Model 2[b]	Model 3[c]
Sharing injecting material	**931.42**[d]	933.20	935.01	**934.54**[d]	936.22	929.84
Share sniffing material	1018.34	**1013.87**[d]	1015.21	1022.81	**1019.41**[d]	1021.32
Ever IDU	**1264.33**[d]	1265.33	1267.20	**1267.31**[d]	1267.85	1269.85
Current IDU	1264.35	**1263.48**[d]	1265.16	1267.54	**1267.29**[d]	1269.27
Ever been to Prison	1312.46	**1309.73**[d]	1311.43	1321.11	**1320.34**[d]	1322.34
Homosexual(MSM)	**981.00**[d]	983.00	983.37	**988.14**[d]	990.14	991.61
Gender	1523.67	**1520.90**[c]	1522.45	1531.73	**1530.35**[d]	1532.35
Educational level	1485.74	**1484.42**[d]	1484.52	1490.91	**1490.77**[d]	1492.19
HIV status	**1517.90**[d]	1518.80	1518.63	**1526.80**[d]	1527.24	1528.75
Center	**1514.75**[d]	1517.04	1517.82	**1523.12**[d]	1526.91	1514.17
Age at test	1503.58	**1503.09**[d]	1505.06	1505.54	**1504.79**[d]	1528.71
Log duration of exposure	**446.49**[d]	447.33		**445.76**[d]	447.03	

[a] Constant odds ratio model
[b] Model with odds ratio depending on the risk factors
[c] Model with odds ratio depending on the risk factors and log duration of exposure
[d] Best model

Table 6: Parameter estimates and standard errors for the association for the best fitting BDM models

Model with log odds ratio depending on the risk factors		
	α_0	α_1
Gender	1.21(0.27)	1.71(1.73)
Sharing sniffing materials	0.82(0.52)	1.87(0.71)
Current IDU	0.93(0.37)	1.06(0.65)
Ever been to prison	2.64(0.82)	-1.75(0.86)
Age at test	3.90(1.60)	-0.08(0.05)
Educational level	1.13(0.29)	1.08(0.64)
Constant log odds ratio models		
Sharing injecting materials	1.26(0.34)	
Ever IDU	1.24(0.29)	
Homosexual(MSM)	1.16(0.32)	
HIV status	1.44(0.26)	
Center	1.41(0.26)	

Table 7: Parameter estimates for the association measures across the levels of the behavioral risk factors

Factors		BDM	ALR	BPM
Gender	Female	2.92 (1.057,4.78)	2.89(1.42,4.37)	1.66(0.75,2.57)
	Male	1.21(0.68,1.75)	1.21(0.64,1.78)	0.82(0.47,1.17)
Age at test		-0.08(-0.17,0.02)	1.38(0.85,1.90)	-0.03(-0.09,0.04)
Educational level	Higher	2.20(1.09,3.32)	2.38(1.14,3.62)	1.32(0.64,2.01)
	Lower	1.13(0.56,1.70)	1.15(0.56,1.74)	0.76(0.40,1.12)
Sharing sniffing materials	Yes	2.69(1.74,3.69)	2.70(1.63,3.77)	1.62(1.07,2.16)
	No	0.82(-0.20,1.85)	0.8(-0.11,1.83)	0.55(-0.11,1.22)
Current IDU	Yes	1.98(0.94,3.02)	2.05(1.21,2.89)	1.18(0.63,1.74)
	No	0.92(0.19,1.66)	0.99(0.24,1.74)	0.62(0.15,1.10)
Ever been to Prison	Yes	0.89(0.33,1.45)	0.91(0.31,1.50)	0.63(0.26,1.00)
	No	2.64(1.04,4.24)	2.92(1.60,4.24)	1.38(0.55,2.22)

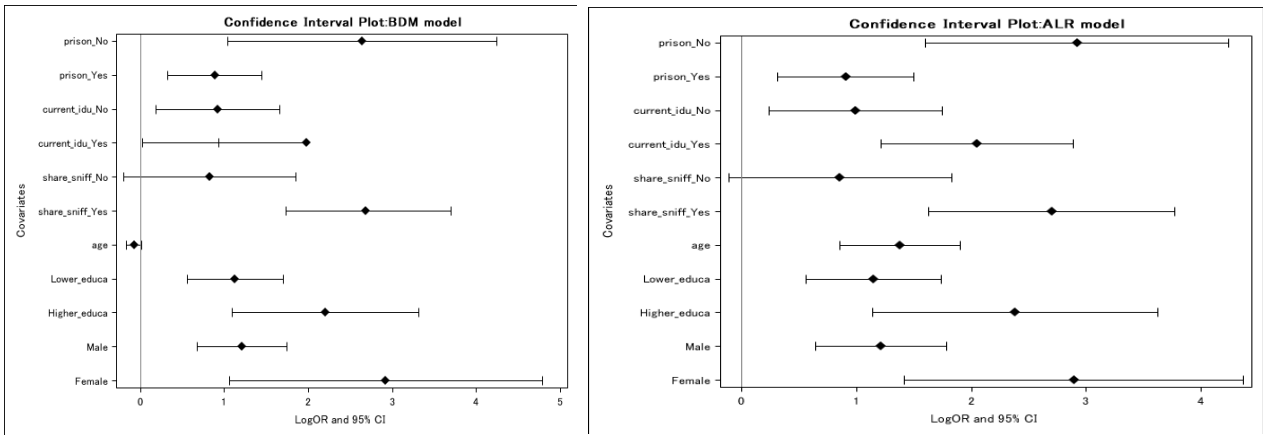


Figure 4: Confidence interval plot for BDM and ALR models for the models depending on the risk factors

The observed and predicted values for the BDM models fitted above are shown in Appendix Figure 1 and it can be seen that the models fitted the data well.

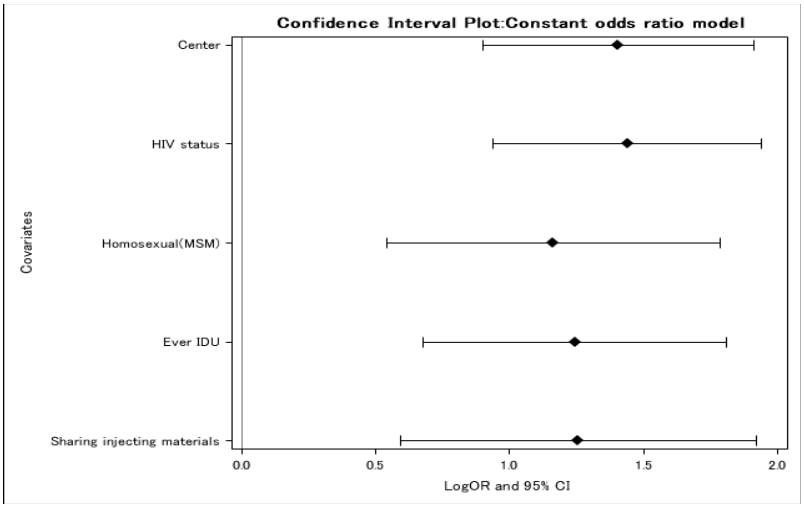


Figure 5: Confidence interval plot for BDM models for the constants odds ratio model

4.2 Shared Random effects model

To investigate the heterogeneity in acquiring infections between one individual to another,shared random effects model were fitted with both logit , cloglog and probit link functions to test for the significance of the σ_b^2. The results of the basic model that includes log duration of drug use only as a covariate are shown in Table 4. Based on AIC values, a model with cloglog link function had the least AIC value, however for comparison purposes with the marginal models presented above, the logit link was taken. The estimate of the variance of the random patient intercept is 2.0943(CI:1.0582,3.5913) and the estimated standard error of this variance component was estimated to be 0.7763. The parameter estimates for the extended models are presented in Table 8 . It can be shown that the variability is significant in all the cases apart from the homosexuals(MSM) which had a borderline significance. There appears to be significant patient-to-patient variation in the way of acquiring infection and the infections are correlated at the individual level. It can also be shown that the variability is highest for the those who were sharing sniffing materials in the last 12 months as compared to other risk factors.

Table 8: Parameter estimates for the shared random effect model

Risk factor	HBV(s.e)	HCV(s.e)	σ^2
Sharing injecting materials	0.34(0.38)	1.26(0.31)*	1.79(0.62,3.70)
Sharing sniffing materials	-0.59(0.42)	-0.74(0.32)*	3.59(1.79, 6.56)
Ever idu	-1.20(0.40)*	-3.56(0.43)*	1.71(0.73, 3.20)
Current idu	0.75(0.30)*	2.19(0.28)*	1.90(0.84, 3.52)
Gender	0.09(0.38)	-0.59(0.26)*	2.13(1.08, 3.65)
HIV status	0.20(0.81)	2.45(0.78)*	2.07(1.04, 3.55)
Homosexual(MSM)	1.29(0.16)	-1.47(0.20)	0.67(0.00, 1.99)
Prison	1.33(0.40)*	0.96(0.24)*	1.41(0.50, 2.76)
Age at test	-0.12(0.03)*	-0.14(0.03)*	2.15(1.11,3.64)

* p-value$<$0.05

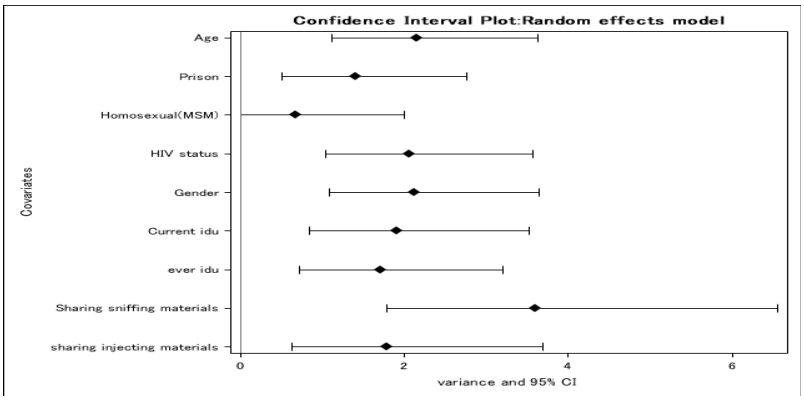


Figure 6: Confidence interval plot for the variance

5 Discussion

HBV and HVC still remains a public health concern among drug users due to their nature of transmission. In this study serological data was collected on individual drug users who had access to drug treatment centers and prisons who agreed to participate in the sero-behavioral study. In this analysis more than one disease were of interest resulting into a clustered data within an individual. Specifically HBV and HCV infections were considered at individual level.

The objective of this study was to analyse the co-infection of HBV and HCV infection and their association on the risk factors. Marginal and cluster specific statistical models for the analysis of clustered data were considered for this analysis. For the marginal models, both fully likelihood(BDM and BPM) and quasi-likelihood methods(ALR) using different link functions were considered. The bivariate models not only improve efficiency but also allow us to study the association between infections as well their dependence on covariates. Results from the models confirm that there is significant association between HBV and HCV. This may be due to the shared mode of transmission for the two infections; getting in contact with infected blood or blood products.

The study has also demonstrated that gender ,educational level, sharing of sniffing materials, being a current IDU and ever been to prison seems to be a major factor in the spread of both HBV and HCV infections among drug users. Sharing of injecting materials was not identified as a risk factor for the prevalence of the co-infection of HBV and HCV , however it has an impact on the spread of the infection in that those who have shared injecting materials had higher odds of getting co-infected compared to those who have never shared injecting materials. This finding was consistent with the study of joint analysis of the co-infection of HCV and HIV infections [2]. Age at the time of interview was found not to be significant ,but the probability of the co-infection decrease with an increase with age. This may suggests that younger population who are drug users are at risk of getting co-infected than older population who are using drugs and campaign to fight illegal drug use should be targeted to this younger population.

Random effects models were fitted to study subjects heterogeneity in acquiring infections. This models has a conditional interpretation unlike the marginal models that had population average interpretations. Considering the variability, there was a significant variability for all the variables apart from homosexuals(MSM) that had borderline significance. Some behavioral risk factors showed higher variability compared to others with the highest variability shown among those who were sharing sniffing materials in the last 12 months, showing its significant contribution to the co-infection. Other behavioral risk factors such as current IDU, gender, HIV status and age at test showed variability in acquiring infections. The heterogeneity shown within the individuals may be due to unmeasured shared genetic and environmental factors

experienced by an individual (cluster).

References

[1] Anon. Hepatitis C: global prevalence. *Weekly Epidemiologica Record*,72(46):341-348

[2] Del Fava, E., Shkedy, Z., Aerts, M., Suligoi, B., Camoni, L., Vallejo, F., Wiessing, L., Kretzschmar, M. (2011). Joint modeling of HCV and HIV Co-Infection among Injecting Drug Users in Italy and Spain Uisng Individual Cross-sectional Data. Available at: http://doclib.uhasselt.be/dspace/bitstream/1942/12288/1/DelFava2.pdf. Accessed on July 07/2012.

[3] Rantala, M., Van de Laar, M. (2008). Surveillance and Epidemiology of Hepatitis B and C in Europe. Available at: http://www.eurosurveillance.org/images/dynamic/EE/V13N21/art18880.pdf. Accessed on July 07/2012.

[4] Mathei C., Shkedy Z., Denis B., Aerts M., Molenberghs G., Van Damme P., Buntinx, F. (2011). Evidence for a substantial role of sharing of injecting paraphernalia other than syringes/needles to the spread of hepatitis C among injecting drug users. *J Viral Hepat*, **9**:157-173.

[5] European Monitoring Center for Drugs and Drug addiction.Viral hepatitis and drug use. Available at http://www.emcdda.europa.eu/topics/hepatitis. Accessed on July 08/2012.

[6] Carey, V., Zeger, SL., Diggle, P. (1993). Modelling multivariate binary data with alternating logistic regression. *Biometrica*, **80**:517-526.

[7] World Journal of Hepatology (2010). Seroprevalence of HCV and its co-infection with HBV and HIV among liver disease patients of South Tamil Nadu. *World Journal of Hepatology*. Available at: http://http://www.wjgnet.com/1948-5182/pdf/v2/i1/42.pdf. Accessed on July 15/2012.

[8] Shobha, S. R. and Cherukuri, M. (2006). Management of Hip Fracture: The Family Physician's Role. Available at: http://www.crediblemed.com/Hip_Fract_Manage.pdf. Accessed on April 02/2012.

[9] Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. New York: Springer.

[10] Diggle P., Heagerty P.,Liang K and Zeger S. (2002). *Analysis of Longitudinal Data*. New York: Oxford university press inc.

[11] Verbeke G., Molenberghs G. and Rizopoulos, D.(2010). *Longitudinal Research with Latent variables*..New York:Springer.

[12] Desenclos, JC. (2003). The challenges of Hepatitis surveillance in Europe. Available at: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=409. Accessed on July 02/2012.

[13] M. Beutels, P. Van Damme, W., Aelvoet, J.Desmyter, F.Dondeyne, C.Goilav, R.Mak, L.Muylle, D.Pierrd, A.Strobant, F.Van Loock, P.Waumans, R.Vranckx (1997). Prevalence of hepatitis A,B and C in the Flemish population *European Journal of Epidemiology* **13**:275-280.

[14] Hou J., Liu Z., Gu F. (2005). Epidemiology and Prevention of Hepatitis B Virus Infection *International Journal of Medical Sciences* **2(1)**:50-57.

[15] Mathei C , Robaeys G, Van Damme P, Buntinx F, Verrando R (2005).Prevalence of hepatitis c in drug users in Flanders:determinants and geographic differences *Epidemiol Infect.* **133(1)**:127-36.

[16] Denis B , Dedobbeleer M, Collet T, Petit J, Jamoulle M, Hayani A,Brenard R (2000). High Prevalence of hepatitis c infection in Belgian intravenous drug users and potential role of the "cotton-filter" in transmission:the GEMT Study *Acta Gastroenterol Belg.* **63(2)**:147-56.

[17] Li-Ping Deng , Xi-En Gui, Yong-Xi Zhang, Shi-Cheng Gao, Rong-Rong Yang (2009).Impact of human immmunodeficiency virus infection on the course of hepatitis C virus Infectcion: A meta-analysis *World J Gastroenterol* **15(8)**:996-1003.

[18] Lars Peter Skovgaard Larsen (2010).Role of contrast enhanced ultrasonography in the assessment of hepatic metastases:A review *World Journal of Hepatology* **2(1)**:8-15.

[19] Center for Disease Control and Prevention(2011).HIV and Viral Hepatitis. Available at: http://www.aids.gov/pdf/hiv-viral-hepatitisacc.pdf. Accesed on 4/9/2012.

[20] Carey C. V., Zeger S. L. and Diggle P. (1993). *Modelling multivariate binary data with alternating logistic regressions.*.Great Britain: Biometrika

[21] Stokes M. E., Davis C. S. and Koch G. G. (2000). *Categorical Data Analysis using the SAS System.*.Cary NC: SAS Institute Inc

[22] Kung-Yee Liang, Scott L. Zeger, Bahjat Qaqush (1992).Multivariate Regression Analysis for Categorical Data. *Journal of the Royal statistical Society* **54(1)**:3-42.

[23] Verbeke G. and Lesaffre E. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association*, 91:217-221.

[24] Molenberghs G. and Leasaffre E.(1994). Marginal modeling of correlated ordinal data using a multivariate plackett distribution. *Journal of the American Statistical Association*, 89:633-644.

[25] Goodman L. A. . Association models and canonical correlation in the analysis of cross-classifications having ordered categories. *Journal of the American Statistical Association*, 76(1981), 320-334.

[26] Bishop Y. M. M. , Fienberg S. E. and Holland P. W. . *Discrete Multivariate Analysis: Theory and Practice.*. Cambridge: MIT Press.

[27] Diamond L.D. and McDonald J.M. . *Demographic Application of Event History.*Oxford University Press

[28] Keiding,N., Begtrup,K., Scheike, G., and Hasibeder, G. (1996). Esimation from current status data in continuos time,*Lifetime data analysis*,**2**:119-129

[29] McCullagh P. and Nelder J.A.(1989). *Generalized Linear Models.*Cambridge University Press

[30] Dale J. R.(1986) Global cross-ratio models for bivariate, discrete, ordered responses, *Biometrics*, 42:909-917.

[31] Hens N.,Shkedy Z., Aerts M., Faes C.,Van Damme P.,Beutels P (2010). *Modelling Infectious Disease Parameters Based on Serological and Social Contact Data. A Modern Statistical Perspective*: Springer.

6 Appendix

Descriptive analysis of the risk factors

The influence of the risk factors prevalence on the infections was explored using multiple logistic regression model where presence/absence of each of infections were considered as the outcome and the risk factors identified by both the chi-square test of independent (Appendix Table 1) as the covariates in the model. First for HBV infections, a model was fitted using logit link functions. An attempt to add interaction terms to the model showed that none of the interactions were significant and were therefore dropped from the model. In Appendix Table 3, it shows that education level, ever vaccinated against HBV, ever been to prison and age at the time of interview were significant risk factors for HBV infection. The odds of getting infected by HBV for a drug user who had higher educational level was estimated to be 0.3045[95% C.I:0.1043,0.8891] times lower than the odds for a drug user who had a lower education level. Those who had received vaccine for HBV had 0.3188[95% C.I: 0.1194,0.8513] times lower odds of getting infection compared to those who have never been vaccinated. The estimated odds of getting infected by HBV for drug users who have ever been to prisons were estimated to be 3.2143[95% C.I: 1.0717,9.608] times higher compared to those who have never been to prison and finally, for a unit increase in age, the odds of getting infected with HBV increase 1.1510[95% C.I: 1.0179,1.3013] times.

Similarly, for the HCV infection, a model with and without interaction were fitted (Appendix Tables 3) . The AIC value for the logit link for the interaction model (AIC=225.65) was smaller than the one for the additive model (AIC=236.38. The final model is presented in Appendix Table F.2 . Gender, vaccination against HBV, age at the time of interview and second order interactions(sharing injecting materials in the last 12 months x being a current IDU, educational level x being in prison and duration of drug use x being in prison) were found to be significant risk factors for HCV infection. The estimated odds of a female drug user getting infected with HCV were found to be 4.24[95% C.I= 4.24,12.43] times higher compared to male drug users, those who have been estimated against HBV had lower odds of getting infected with HCV compared to those who had not been vaccinated [OR=0.2633; 95% C.I=0.1141,0.6076]. HCV infection was also found to be increasing with age of the patients [OR=1.2078;95% C.I=1.0446,1.3967]. The prevalence of HCV for those who have been in prison and those who have never been to prison depends on the duration of drug use. The odds of getting infected with HCV infection for those who are sharing injecting material depends on whether they are current injectors or not.

Table F.1: Demographic and behavioral characteristics for HBV and HBV status according to risk factors

Variable	HBV +(%)	HBV -(%)	HBV χ^2 p-value	HCV +(%)	HCV -(%)	HCV χ^2 p-value
Gender (N=972)			0.3442			0.3055
Male	77(83.70)	700(79.55)		222(77.89)	555(80.79)	
Female	15(16.30)	180(20.45)		63(22.11)	132(19.21)	
Education level(N=961)			0.0027			< .0001
High education	73(80.22)	562(64.60)		219(78.21)	416(61.09)	
Low education	18(19.78)	308(35.40)		61(21.79)	265(38.91)	
Center type(N=972)			0.0005			< .0001
AC & MSOC	67(72.83)	470(53.41)		198(69.47)	339(49.34)	
CGG,TG & WGC	11(11.96)	120(13.64)		35(12.28)	96(13.97)	
CIC	3(3.26)	167(18.98)		23(8.07)	147(21.40)	
PAAZ & PH	11(11.96)	123(13.98)		29(10.18)	105(15.28)	
Ever shared injecting material (N=467)			0.1485			< .0001
Yes	54(76.06)	267(67.42)		193(78.78)	128(57.66)	
No	17(23.94)	129(32.58)		52(21.22)	94(42.34)	
Ever shared sniffing material (N=704)			0.0523			0.0018
Yes	38(63.33)	482(74.84)		127(65.46)	393(77.06)	
No	22(36.67)	162(25.16)		67(34.54)	117(22.94)	
HIV test results(N=972)			0.3416			< .0001
Positive	3(3.26)	16 (1.82)		15(5.26)	4(0.58)	
Negative	89(96.74)	864(98.18)		270(94.74)	683(99.42)	
Vaccination against HBV (N=558)			0.0016			0.0002
Yes	11(21.15)	222(43.87)		57(30.65)	176(47.31)	
No	41(78.85)	284(56.13)		129(68.35)	196(52.69)	
Current injecting drug users(N=798)			0.0023			< .0001
Yes	47(55.29)	272(38.15)		173(64.55)	146(27.55)	
No	38(44.71)	441(61.85)		95(35.45)	384(72.45)	
Ever-IDU(N=821)			< .0001			< .0001
Yes	78(88.64)	468(63.85)		269(96.07)	277(51.01)	
No	10(11.36)	265(36.15)		11(3.89)	264(48.80)	
Ever been to prison(N=794)			<.0001			<.0001
Yes	72(87.80)	411(57.72)		206(78.63)	277(52.07)	
No	10(12.20)	301(42.28)		56(21.37)	255(47.93)	
MSM (N=484)			0.2635*			0.1455*
Yes	2(4.55)	9(2.05)		1(0.72)	10(2.90)	
No	42(95.45)	431(97.95)		138(99.28)	335(97.10)	
Duration of drug use			< .0001			< .0001
Age at test			< .0001			< .0001

* Fisher's exact p-value

Table F.2: Parameters estimates and C.I for the Logit links models for HBV infection

	Logit
Parameter	**Estimate(C.I)**
Intercept	-7.7063(-11.4385,-3.9741)*
Females	0.6248(-0.3372,1.5867)
Higher education level	-1.300(-2.4606,-0.1394)*
Center 1	0.7231(-0.4803,1.9265)
Center 2	-0.7492(-3.0937,1.5954)
Center 3	-1.2016(-3.5143,1.1112)
Sharing sniffing materials	-0.2868(-1.1022,0.5286)
Vaccinated against hbv	-1.2216(-2.2829,-0.1602)*
Current idu	0.1178(-0.7622,0.9977)
Ever idu	1.2597(-0.2326,2.7521)
Ever been to prison	1.2955(0.1260,2.4650)*
Age at interview	0.1428(0.0048,0.2808)*
Duration of drug use	-0.0564(-0.1895,0.0767)
AIC	193.6887

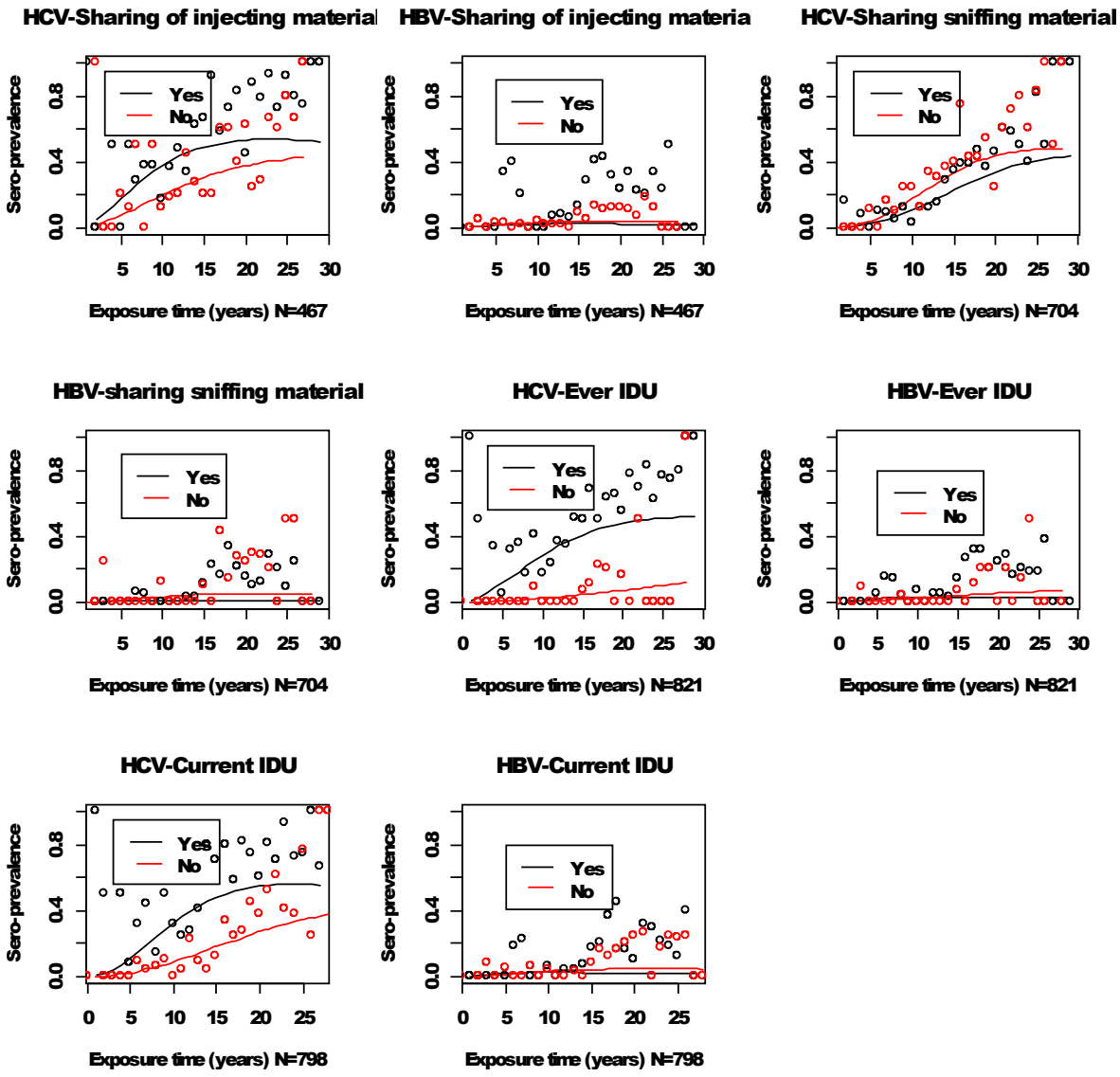
* p-values < 0.05

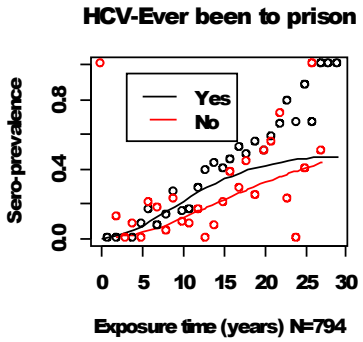
Table F.3: Parameters estimates and C.I for the additive and interaction models for HCV infection-logit link

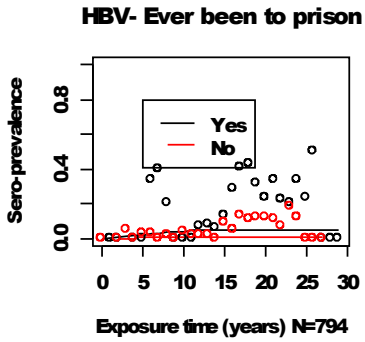
	Additive model	Model with interaction
Parameter	**Estimate(95% C.I)**	**Estimate(95% C.I)**
Intercept	-6.2871(-9.2958,-3.2783)	-5.8564(-9.2930,-2.4199)*
Females	1.3811(0.4084,2.3539)	1.4824(0.4448,2.5199)*
Higher education level	-1.2651(-2.0582,-0.4720)	-2.4971(-3.8742,-1.1199)*
Center 1	0.6339(-0.4245,1.6917)	0.2701(-0.8791,1.4192)
Center 2	-1.1386(-2.5942,0.3171)	-1.4978(-3.0366,0.0409)
Center 3	-0.5231(-1.9565,0.9103)	-1.1483(-2.6864,0.3898)
Sharing injecting materials	1.3056(0.5502,2.0609)	2.6484(1.2892,4.0077)*
Sharing sniffing materials	-2207(-0.9451,0.5038)	-0.0368(-0.8045,0.7309)
Vaccinated against hbv	-0.2175(-0.3370,-0.0979)	-1.3344(-2.1706,-0.4982)*
Current idu	0.1984(0.0841,0.3126)	2.6595(1.2625,4.0566)*
Ever been to prison	1.2338(0.4170,2.0506)	-1.5254(-3.7328,0.6819)
Age at interview	0.1735(0.0355,0.3115)	0.1888(0.0436,0.3341)*
Duration of drug use	-0.0615(-0.1956,0.0725)	-0.1498(-0.3036,-0.0040)*
shareinj*Currentidu		-1.9793(-3.6341,-0.3245)*
Educlevel*prison		2.2371(0.5332,3.9411)*
durationdrug*prison		0.1436(0.0117,0.2756)*
AIC	236.375	225.6481

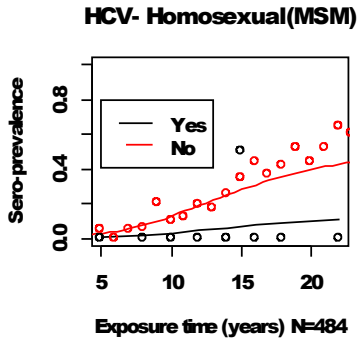
* p-values < 0.05

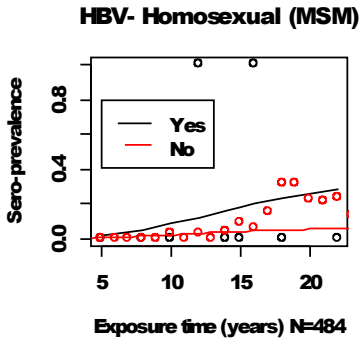
Fitted and observed graphs for the final model

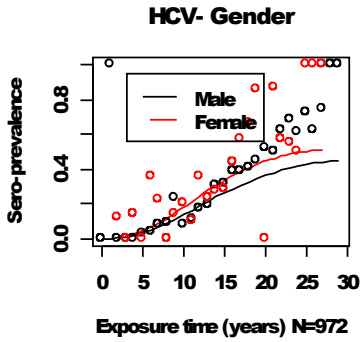


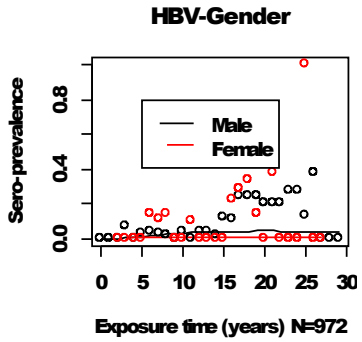

HCV-Ever been to prison
Sero-prevalence
0.0
0.4
0.8
Yes
No
0
5
10
15
20
25
30
Exposure time (years) N=794

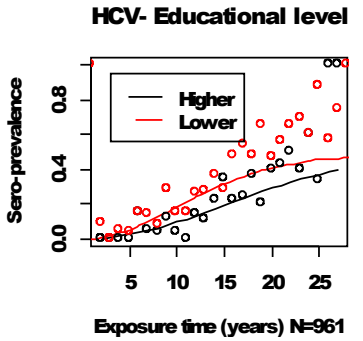

HBV- Ever been to prison
Sero-prevalence
0.0
0.4
0.8
Yes
No
0
5
10
15
20
25
30
Exposure time (years) N=794

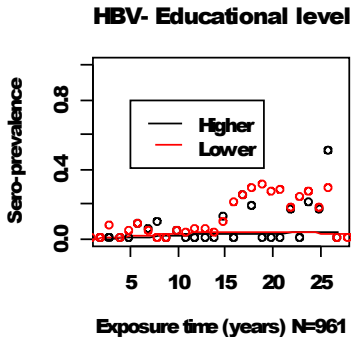

HCV- Homosexual(MSM)
Sero-prevalence
0.0
0.4
0.8
Yes
No
5
10
15
20
Exposure time (years) N=484

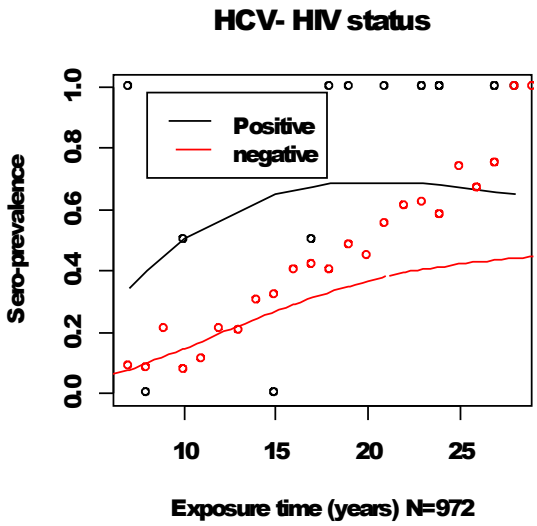

HBV- Homosexual (MSM)
Sero-prevalence
0.0
0.4
0.8
Yes
No
5
10
15
20
Exposure time (years) N=484

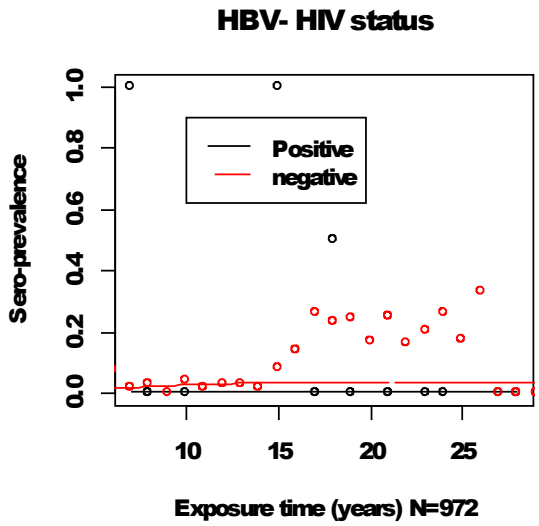

HCV- Gender
Sero-prevalence
0.0
0.4
0.8
Male
Female
0
5
10
15
20
25
30
Exposure time (years) N=972

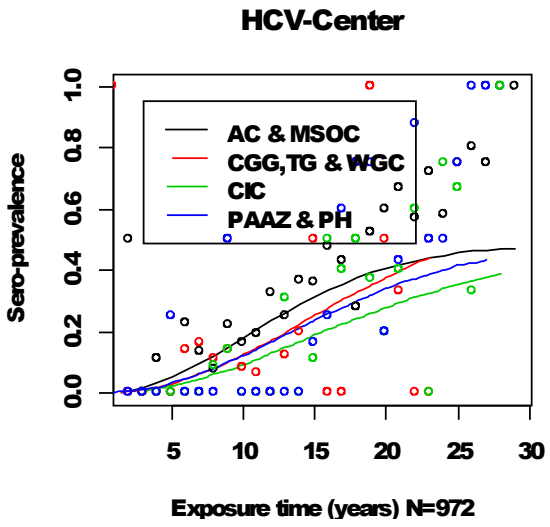

HBV-Gender
Sero-prevalence
0.0
0.4
0.8
Male
Female
0
5
10
15
20
25
30
Exposure time (years) N=972


HCV- Educational level
Sero-prevalence
0.0
0.4
0.8
Higher
Lower
5
10
15
20
25
Exposure time (years) N=961


HBV- Educational level
Sero-prevalence
0.0
0.4
0.8
Higher
Lower
5
10
15
20
25
Exposure time (years) N=961


HCV- HIV status
Positive
negative
Sero-prevalence
0.0
0.2
0.4
0.6
0.8
1.0
10
15
20
25
Exposure time (years) N=972


HBV- HIV status
Positive
negative
Sero-prevalence
0.0
0.2
0.4
0.6
0.8
1.0
10
15
20
25
Exposure time (years) N=972


HCV-Center
AC & MSOC
CGG,TG & WGC
CIC
PAAZ & PH
Sero-prevalence
0.0
0.2
0.4
0.6
0.8
1.0
5
10
15
20
25
30
Exposure time (years) N=972

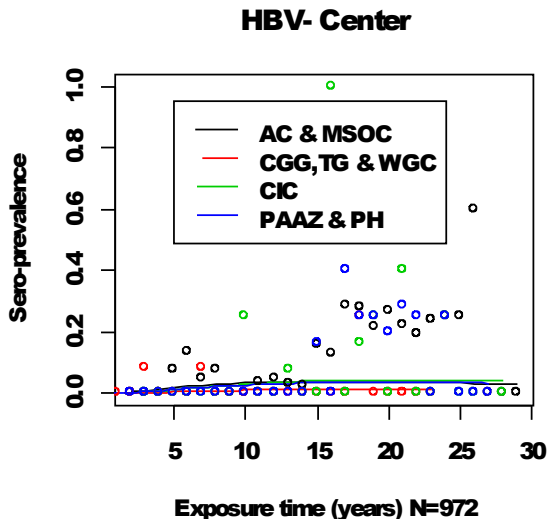

HBV- Center
AC & MSOC
CGG,TG & WGC
CIC
PAAZ & PH
Sero-prevalence
0.0
0.2
0.4
0.6
0.8
1.0
5
10
15
20
25
30
Exposure time (years) N=972

Table F.4: Variables description

Variable	Format	Label
id		Unique identification number
center		Type of center
sex		Gender
	1=male	
	2=female	
hcv		HCV test results
	0=negative	
	1=positive	
hbv		HBV test results
	0=negative	
	1=positive	
hiv		HIV test results
	0=negative	
	1=positive	
age		age (yrs) at the time of interview
vaccinhepb		Vaccinated against HBV
	0=no	
	1=yes	
durainject		duration(yrs) of injecting drug career
Everidu		Ever Injecting drug
	0=no	
	1=yes	
Currentidu		Current injecting drug users
	0=no	
	1=yes	
durationdrug		duration(yrs) drug of use
Educlevel	EDUCATION	highest Education level attained
homosexual		homosexual
	0=no	
	1=yes	
prison		Ever been to prison
	0=no	
	1=yes	
shareinj		sharing injecting materials in the last 12 months
	0=no	
	1=yes	
sharesniffm		sharing sniffing materials in the last 12 months
	0=no	
	1=yes	

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
Joint modelling of HBV and HCV infections from cross-sectional serological data

Richting: **Master of Statistics-Biostatistics**
Jaar: **2012**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

Orwa, James

Datum: **14/09/2012**