2012
2013

FACULTY OF SCIENCES

Master of Statistics: Biostatistics

Masterproef

Estimate prevalence of Hepatitis B and C among Injecting Drug Users and identify risk factor associated with the infections

Promotor :
Prof. dr. Ziv SHKEDY

Promotor :
Dr. KAATJE BOLLAERTS

Henry Ntanda

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

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

Maastricht University

universiteit hasselt
KNOWLEDGE IN ACTION

2012
2013

FACULTY OF SCIENCES

Master of Statistics: Biostatistics

Masterproef

Estimate prevalence of Hepatitis B and C among Injecting Drug Users and identify risk factor associated with the infections

Promotor :
Prof. dr. Ziv SHKEDY

Promotor :
Dr. KAATJE BOLLAERTS

Henry Ntanda

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

Maastricht University

universiteit hasselt
KNOWLEDGE IN ACTION

Estimating Prevalence of Hepatitis B and C Among Injecting Drug Users and Identifying Risk Factors Associated with the Infections.

BY:

Henry Nicholas Ntanda.

Supervisors:

Prof. dr. Ziv Shkedy.

Prof. dr. Geert Molenberghs.

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

October, 2012

Acknowledgements

Successful completion of this thesis has been through the support of a number of individuals. First of all, I acknowledge my supervisors; Prof. dr. Ziv Shkedy and Prof. dr. Geert Molenberghs who continually guided me through this work, I learnt a lot from you, and also thank you for your time that you dedicated on my summer project. This work would not have been possible without Scientific Institute of Public Health who allowed me to use their data set.

I would like to extend my gratitude to Mr.Yusuf Sembatya and Miss. Sylvia Namugeme Ssentongo who without them I would not be here.

I am grateful to my parents for the support and love they've given me throughout the years. To my daughter Chanelle Namilo, brothers and sisters I am sorry for spending most of the time away from you, and I am sorry for missing you introduction ceremonies, wedding and all the events that I could not attend but always remember your brother loves you so much.

Most of all, I thank God for giving me wisdom, strength and life to successfully complete this program.

Henry Nicholas Ntanda
University of Hasselt
Belgium, October, 2012

Abstract

Background: Drug users especially those injecting intravenously, are at an increased risk of infection with blood-borne viruses including hepatitis B virus (HBV) and hepatitis C virus (HCV). Hepatitis C has emerged as a major threat to public health world-wide, about 170 million people are chronically infected with virus (Anon, 1997). HBV infection is widely present with approximately one third of the world's population has been exposed to the virus, and an estimated 350 million people are chronically infected (WHO 2004). Most HBV infections occur in well defined high risk groups, including drug users (Hou et al., 2005).

Objectives: The aim of this study was to estimate the prevalence of Hepatitis B and C among Injecting Drug Users (IDUs) and to identify risk factors associated with the infections.

Methodology: The data used is from a cross-sectional study (2004 - 2005) carried out to assess drug related infectious diseases in Belgian treatment centres and prisons. A total of 226 treatment centres and 15 prisons were identified and asked to participate. In this study, participation was on a voluntary basis and all treatment centres willing to participate could do so all over the country. The data used contains 979 drug users who participated in the study infected with HCV or HBV. A total of 92 (9.4%), 288 (23.3%) were sero-positive and 887 (90.6%), 691 (70.6%) were sero-negative for HBV and HCV respectively. Drug users included in the study were aged 15 to 40 years. Generalized additive models are used because of their flexibility by using smoothing functions instead of parametric to estimate the overall trend of the data.

Results: Drug users with HCV have a high prevalence compared to HBV. Drug users who have ever injected drugs before, current injecting drugs, homosexuals, being in prison, not vaccinated for HBV, sharing injections, sharing sniffing materials, being HIV positive, low level of education, being male, and being in a particular treatment center were at a high risk of contracting any of the infections the longer they are exposed to drugs.

Keywords: Hepatitis B virus (HBV), Hepatitis C virus (HCV), Drug Users (DUs), Generalized Additive Models (GAM).

Certification

This is to certify that this report was written by **Henry Nicholas Ntanda** under our supervision.

Signature:..... Date.....

Prof. dr. Ziv Shkedy. **Internal Supervisor.**

Signature:..... Date.....

Prof. dr. Geert Molenberghs. **External Supervisor.**

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

Signature:..... Date.....

Henry Nicholas Ntanda. **Student.**

Contents

1 Introduction

Injection drug users, especially those injecting intravenously, are at an increased risk of infection with blood-borne viruses, including hepatitis B virus (HBV), hepatitis C virus (HCV) and human immunodeficiency virus (HIV). Hepatitis C is an infectious disease affecting the liver, caused by the hepatitis C virus. There is no vaccine against HCV available, due to the extensive genetic heterogeneity of the virus. The main HCV transmission routes are blood transfusions from unscreened donors, injecting drug use, unsafe therapeutic injections, and other health-care-related procedures (Baker, 2002). The exposure to infected blood in the context of injecting drug use is the predominant way of transmission in the developed countries. (Alter, 2006). HCV infection seems to be acquired rapidly after the initiation of an injecting career and many people may have been infected as a result of occasional experimentation with illicit drugs (Mathei et al., 2006). Injecting drug users are now the group at high risk of the infection accounting for up to 60 - 90% of new infections. In all European Union countries the incidence of HCV among IDUs is extremely high, ranging from about 30% to over 90% (Wiessing et al., 2003). The available data from treatment programmes indicate that the prevalence of HCV infection in IDUs aged under 25 varies from 12% in Tampere, Finland (2001), to around 60% in Dublin, Ireland (1997), and Italy (1999). In Western Europe the prevalence of HCV in the general population is low although it increases from about 0.1% in the North to more than 1% in the South (Desenclos, 2004).

Therefore, over the past few years Hepatitis C emerged as a major threat to public health world-wide. The estimated prevalence of HCV infection worldwide is 2%, representing 123 million people (Perz et al., 2006). In the general population, HCV infection prevalence is respectively equal to 0.5% and 2% (WHO, 1999), while among injecting drug users (IDUs) it is equal to 59.2% and 59.1% - 73.3% (EMCDDA, 2010), respectively.

In Belgium, a prevalence study undertaken in the mid-nineties in the hospitalised population showed anti-HCV in 0.87% of the serum samples (Hutse and Quoilin 2004). A recent study in the general population where analysis was performed on saliva resulted in HCV prevalence of

0.11%. The HCV prevalence in the Belgian general population is therefore likely to be between these two figures. In 2002, HCV prevalence among IDUs asking for treatment was 67% for the French Community (self-reported data collected through the EUROTOX monitoring system). From 2001 to 2004 around 79% of the patients tested at 'Free Clinic' (medical-social low threshold centre situated in the city of Antwerp, average $n = 264$ tested IDUs per year) and 38% of the patients tested at 'De Sleutel' (a Flemish institution composed by several ambulatory and residential treatment centres, average $n = 90$ tested IDUs per year) were positive for Hepatitis C (Sleiman 2005, Raes and Lombaert 2004).

Since then, transmission through contaminated blood products has been effectively prevented through blood screening. Unexplained cases are particularly high among drug users who have no history of injection risk and no other identifiable risk factors. About 12% (Flamm et al., 1998) to 15% (Mcmahon et al., 2004) of the HCV cases cannot be explained by the currently known risk factors. Among those unexplained cases, drug users are particularly high (Flamm et al ., 1998). The high prevalence or incidence figures, low access (Wiessing et al., 2003), the difficult compliance to treatment, and the lack of an effective vaccine underline the need for preventive interventions on the group of young and new injectors. However, the other groups of drug users may not be neglected because primary HCV infection does not confer protective immunity against subsequent infections with viruses of other genotypes. This may also hamper the development of a vaccine (Proust et al., 2000).

Infection with hepatitis B affects the liver and results in a broad spectrum of disease outcomes. This infection can spontaneously resolve and lead to protective immunity, resulting in a chronic infection which in rare cases can cause acute liver failure with a high risk of dying. In contrast to HBV, an infection with HCV becomes chronic in most cases (Lauer and Walker 2001). People with chronic hepatitis B or C virus infection remain infectious to others and are at risk of serious liver disease such as liver cirrhosis or hepatocellular cancer (HCC) later in life (Sorrell et al., 2001). In Western Europe 5 - 7% of the general population is infected, 0.5 - 2% are chronic carriers and most HBV infections occur in well defined high risk groups,

including drug users (Hou et al., 2005). Throughout the EU, approximately 20 - 60% of IDUs have antibodies against HBV. The prevalence of current HBV infection is recorded in only a few countries, but appears to differ widely and is in some cases high (EMCDDA 2003). In Belgium, the HBV prevalence in the hospitalised population (including acute infections, recovered and chronic carriers) is estimated at 7.4% (Beutels et al., 1997). For HbsAg, sero-prevalences of 0.7% (Beutels et al., 1997) and 0.66% (Hutse and Quoilin 2004) were found. The incidence in 1991 - 1992 was 6/100,000 in the general population (Devroey et al., 1994).

In 2002, a study carried out in Belgium estimated the HBV prevalence among IDUs asking for treatment was 9% for the French Community. From 2001 - 2004 on average 57% of the patients tested at 'Free Clinic' (on average 259 IDUs tested per year) and 18% of the patients tested at 'De Sleutel' (on average 65 IDUs tested per year) were positive for Hepatitis B (Sleiman 2005, Raes and Lombaert 2004). While HCV is mainly transmitted through drug injection, unprotected sex is considered the major route of transmission for HBV hence considered to be a sexually transmitted disease (STD) especially in low endemic areas such as Western Europe. However in the United States and Western Europe, injecting drug use remains a very important mode of HBV transmission (23% of all patients) (Hou et al., 2005).

HBV infection is widely present: approximately one third of the world's population has been exposed to the virus, and an estimated 350 million people are chronically infected (WHO 2004). More than 500,000 people die each year of hepatitis B related diseases (WHO 2004). There is a distinct geographical variation in both HBV and HCV prevalence and incidence in the European Union and neighbouring countries.

This study was, therefore, carried out to estimate the prevalence of hepatitis B and C over exposure time among drug users in contact with treatment centres and to identify risk factors associated with the different infections. To achieve the objective of the study, semi-parametric models will be applied with hepatitis B status (positive or negative) and hepatitis C status (positive or negative) as the outcomes of interest.

The thesis is organized as follows: Section 2 provides a description of the data set while the statistical methodologies used to achieve the objectives of the study are explained in Section 3. The results from the statistical methods are presented in Section 4 and Section 5 is devoted to discussion and conclusions.

2 Data Description

The data set used in this report came from a cross-sectional study aimed at assessing drug related infections in Belgian treatment centres and prisons. A treatment centre is defined as one recognised and financed by the authority that takes care of people with drug problems including providing treatment (whether or not for it is specific assignment towards drug users). Since drug use is common in prisons, this specific group was included. Due to lack of inventory of the treatment centres at national level, a list of existing treatment centres in Belgium was constructed for the purpose of this study on the basis of the information available in local sources. The existing treatment centres are divided into in-patient and out-patients. A total of 226 treatment centres and 15 prisons were identified and asked to participate. These included: 141 outpatient centres (8 Medical and Social Reception Centres, 46 specialised outpatient centres, 37 Centres of Mental Health Care, and 50 other outpatient centres) and 76 inpatient centres (9 Crisis Intervention Centres, 13 therapeutical programs, 31 Psychiatric Hospitals, and 23 Psychiatric Units of General Hospitals). The centres had to fulfil a number of conditions which included having a medical doctor inside the centre or at least the possibility to collaborate with a medical doctor and storing the blood samples in a refrigerator (at a temperature of 4°C). The study protocol was sent to the 226 centres requesting them to participate in the study, 65 centres accepted to participate, 67 centres refused and the rest did not respond. The participating centres were categorised as follows: Medical social centres for drug users (MSOC/MASS), Non-residential day care centre (AC), Centres for mental health care (CGG, WGC), Crisis intervention centre (CIC), Therapeutic community (TG), Psychiatric unit within general hospital (PAAZ), Psychiatric hospital (PH). To have enough sample size for analysis, the treatment centres were further grouped into 4 categories that is: AC+MSOC, CIC, WGC+CGG+TG, and PAAZH+PH.

The sampling procedure of this study can be summarised as follows. The Regions and Provinces constitute a first stratification factor (considering the Brussels Capital Region as a particular case). However, those figures cannot be compared with the real treatment offer present in each province. In normal circumstances, the selection of the centres should have been done randomly.

However, in this study participation was on a voluntary basis, all centres willing to participate could participate, and each centre received a financial compensation for their participation. All Drug users who participated in the study were interviewed and information collected included: drug use, risk behaviours, legal problems, infectious diseases, socio-demographic issues, contact with drug and health services, and knowledge or attitudes. Subsequent to the interview, a blood sample was taken from the drug users and part of the serum was tested for HBV and HCV, and the rest for HIV. The data contains 979 drug users who participated in the study infected with HCV or HBV. A total of 92 (9.4%), 288 (23.3%) were sero-positive and 887 (90.6%), 691 (70.6%) were sero-negative for HBV and HCV respectively. Drug users included in the study were aged 15 to 40 years and use or have used regularly one or more of the following substances by any route of administration: opiates, opiate antagonists, cocaine, amphetamines, methadone, buprenorphine. The median age of the participating drug users and length of exposure to drugs was 29 and 13 years, respectively. Among the drug users who agreed to participate in the study, 784 (80.08%) were males and 195 (19.92%) females. See appendix Table 3 and 4 for the risk factors considered in the study.

Figure 1 shows that the proportion of drug users infected with HCV increases with duration of exposure time. While that of the HBV is low for the exposure in the group 0 to 15 years but generally the pattern appears non-monotone.

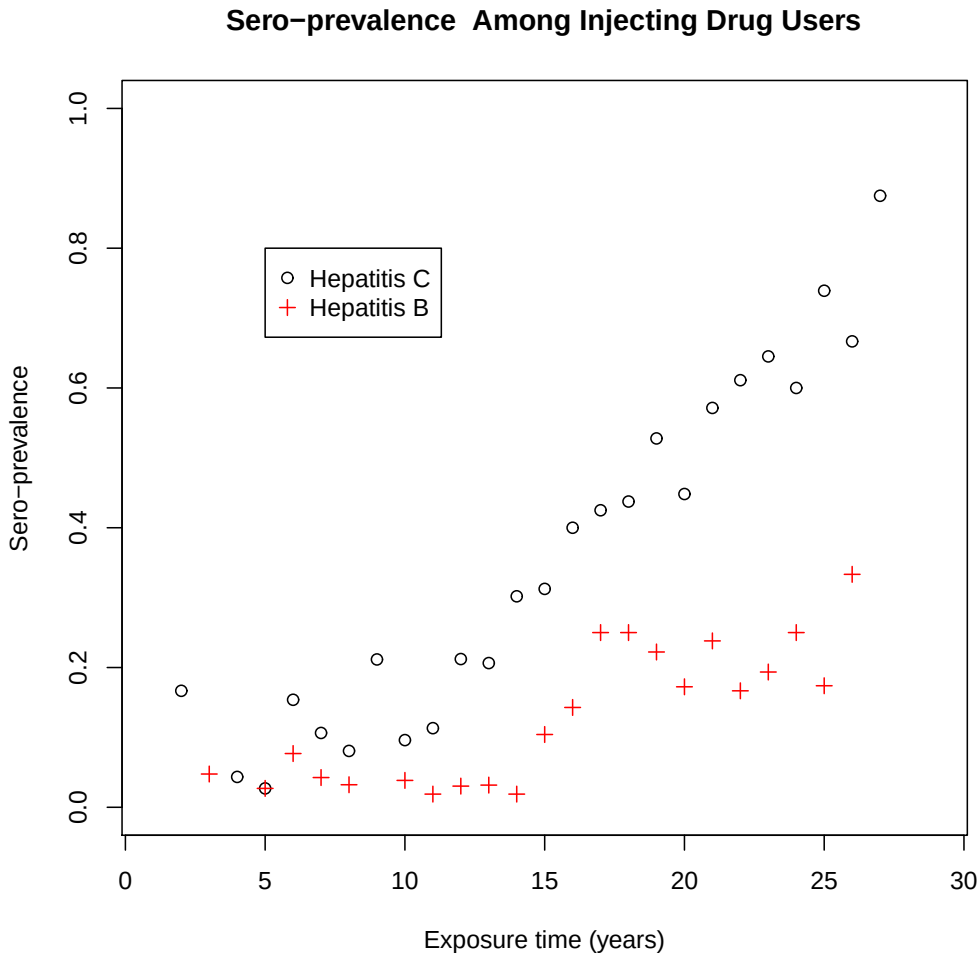


Figure 1: Sero-prevalence of HCV-HBV varying with exposure time in years among IDUs.

3 Methodology

3.1 Generalized linear models

Generalized linear models (Nelder and Wedderburn, 1972) make the distributional assumptions that the response variables y_i are independent follow an exponential family distribution. In this study, the response is binary; that is hepatitis C status (positive versus negative) and Hepatitis B status (positive versus negative).

$$y_i = \begin{cases} 1 & \text{if sero-positive,} \\ 0 & \text{if sero-negative,} \end{cases} \tag{1}$$

y_i is the response that indicates whether an individual i has experienced the infection before exposure time d_i, $(i = 1 \cdot \cdots \cdot n)$. Therefore, it is assumed that $Y_i|d_i \sim$ Bernoulli $\pi(d_i)$, where $\pi(d_i) = Pr(y_i = 1)$ is the probability of being infected with either HCV or HBV. The prevalence is related to exposure time with the model:

$$g(P(Y_i = 1|d_i)) = g(\pi(d_i)) = \eta(d_i), \tag{2}$$

where g is the link function, $P(Y_i = 1|d_i)$ is the mean component, which expresses the probability of being infected given exposure time d_i, and $\eta(d_i)$ is the linear predictor which can consist of risk factors, see appendix Table 3 and 4. In the analysis, one could use generalized linear regression with linear, quadratic, or higher order polynomials in exposure time to capture general trends in the data. The drawback of those models is their inability to capture systematic deviations of the data from the overall trend.

3.2 Generalized additive models (GAM)

A generalised additive model is a generalized linear model with a linear predictor involving a sum of smooth functions of covariates. Generalised additive models are at times preferred over generalized linear models (GLM), as they are considered a viable approach because of their flexibility in using smoothing functions instead of parametric terms to estimate a trend.

Although the GAMs extend GLMs in the same way as additive models extend linear models, a GAM differs from a GLM in the linear predictor (Hastie and Tibshirani, 1986, 1990):

$$g(\mu_i) = X_i^*\theta + f_1(x_{1i}) + f_2(x_{1i}) + f_3(x_{3i}, x_{4i}) + \ldots, \tag{3}$$

where $\mu_i \equiv E(Y_i)$ and Y_i follows some exponential family distribution.
Y_i is a response variable, X_i^* is a row of the model matrix for any strictly parametric model component, θ is the corresponding parameter vector, and the f_j are the smooth functions of the covariates x_k. In a GAM, some variables may enter the additive predictor linearly but the effects of others are modelled as splines. Therefore, GAMs seem to strike a sensible compromise between ease of interpretation and flexibility. A GAM can also contain parametric terms, in similar fashion to the semi-parametric additive model. Therefore this implies that the overall trend for each infection can be represented by an additive model of two components; a linear component Xβ and a smooth component Zu. This is because it is of interest to compare the prevalence curves of the different predictor(s), such as male and females, given the different infections. GAMs were fitted with three different assumptions to estimate the prevalence and to identify the risk factors associated with HBV and HCV, respectively where the responses were considered univariately. Therefore, given a response, the first model that was fitted included a smoothed part of the exposure time (non-parametric), and a risk factors such as gender (x_1) (parametric). Then this model was extended to a second model that included, in addition to the smoothed part (exposure time), an interaction between risk factors gender (x_1) and exposure time (d_i). Finally, since gender (x_1) has two levels, the last model fitted took that into account. The model fitted smoothed exposure time independently by levels . The three procedures were repeated for all risk factors but each risk factor was added to the model, independently of the other. Only smoothed exposure time was held constant. This means that if the first time one considers gender as the risk factor, in the next set of fitted models gender is dropped and another risk factor is considered. The models fitted are given below:

Model 1: This model assumes that the difference amongst the groups, if present, does not depend on exposure time alone. Therefore we assume that the underlying linear trend in the

groups differs by a shift (γ_0) only. The model can be represented as :

$$\eta(d_i) = (\beta_0 + \gamma_0 G_i) + \beta_1 d_i + \sum_{k=1}^{K} \mu_k (d_i - K_k)_+^p, \tag{4}$$

where $k_k, k = 1, ...K$ are knots which together with p determine the smoothness. $\eta(d_i)$ is a linear predictor, the beta's are coefficients of the parametric part and $\sum_{k=1}^{K} \mu_k (d_i - K_k)_+^p$ is the nonparametric part of the model, μ_k are the coefficients of the non-parametric part which are common to all groups. Since fitting a P-spline results in a rough fit, therefore a restriction can be imposed on μ_k by penalising roughness of the fit, hence putting a constraint on $\mu_k \sim N(0, \sigma_\mu^2)$. Finally $(d_i - K_k)_+^p$ is the basis function, which in this case is the cubic spline regression that is used for analysis. Cubic regression splines are part of the general class of regression splines, joining (cubic) polynomials at the knots of the spline to ensure continuity and differentiability up to degree two. The number of knots is smaller than the unique number of data points and the placement of knots is user-defined. The cubic regression spline is available in the R-library "*mgcv*" (Wood 2006). The default knot location is governed by the quantiles of the covariate distribution. The within variability for the coefficients of the non-parametric part for each level of the group is given by $\text{Var}(\mu_k) = \sigma_u^2$ and G_i is a group indicator like gender.

Model 2: The assumption is that the linear part of the model differs, while the same smooth part is considered for all groups. In this case, the group effect is no longer constant over duration. Also, the model can be represented as:

$$\eta(d_i) = (\beta_0 + \gamma_0 G_i) + (\beta_1 + \gamma_1 G_i) d_i + \sum_{k=1}^{K} \mu_k (d_i - K_k)_+^p, \tag{5}$$

with $Var(\mu_k) = \sigma_u^2$.

Model 3: The assumption of a constant smoothing parameter and the coefficients of the non-parametric part across the groups is relaxed, thereby assuming that the groups can be smoothed separately with different smoothing parameters. Hence, both the fixed-effects part and the non-parametric part differ by group but with the variance component $Var(\mu_{kg}) = \sigma_{ug}^2$ being group - specific'.

$$\eta(d_i) = (\beta_0 + \gamma_0 G_i) + (\beta_1 + \gamma_1 G_i) d_i + \sum_{k=1}^{K} \mu_{kg} (d_i - K_k)_+^p. \tag{6}$$

3.3 Penalised splines

Semi-parametric models are extensions of parametric analysis but include segment-wise parametric functions that are able to follow deviations from the overall trend in the data. Therefore, the linear predictor can be estimated semi-parametrically using penalized splines (Ruppert et al., 2003). Taking a p^{th} degree spline model with K knots,

$$\eta(d_i) = \beta_0 + \beta_1 d_i + \cdots + \beta_p d_i^p + \sum_{k=1}^{K} \mu_k (d_i - K_k)_+^p, \tag{7}$$

with truncated power base functions defined as

$$(d_i - K_k)_+^p = \begin{cases} 0 & \text{if } d_i \leq K_k, \\ (d_i - K_k)_+^p & \text{if } d_i > K_k, \end{cases} \tag{8}$$

where $d_1 \leq d_2 \leq \cdots \leq d_N$, denotes the k^{th} knot.

The mean structure for the model $\eta(d_i)$ can be presented in vector form like $\eta = \mathrm{X}\beta + Zu$. We assume that $\eta_i = \eta(d_i)$, therefore $\eta = (\eta(d_i), \cdots, \eta(d_N))^T$, $\beta = (\beta_0, \beta_1, \cdots\cdots, \beta_p)^T$ to represent the vector of the coefficient of the fixed effects; $u = (u_1, u_2, \cdots, u_k)^T$ is the vector of random effects and the design matrices are:

$$X = \begin{bmatrix} 1 & d_1 & d_1^2 \ldots & d_1^p \\ 1 & d_2 & d_2^2 \ldots & d_2^p \\ \vdots & \vdots & \ldots & \vdots \\ 1 & d_N & d_N^2 \ldots & d_N^p \end{bmatrix}, Z = \begin{bmatrix} (d_1 - K_1)_+^p & (d_1 - K_2)_+^p \ldots & (d_1 - K_k)_+^p \\ (d_2 - K_1)_+^p & (d_2 - K_2)_+^p \ldots & (d_2 - K_k)_+^p \\ \vdots & \vdots & \vdots \\ (d_N - K_1)_+^p & (d_N - K_2)_+^p \ldots & (d_N - K_k)_+^p \end{bmatrix}.$$

A large number of knots, between 5 and 20, are considered to attain the desired flexibility. However, this brings the problem of overfitting . Therefore, to overcome this problem, the non-linear part Z is penalised by assuming that the coefficients μ are random effects and are constrained to reduce the influence of the knots and hence to ensure stable estimation. Also it is assumed that $\mu \sim N(0, \sigma_u^2 I)$.

3.4 Estimating the smoothing parameter

Smoothing the data using penalised splines requires choosing the value for the smoothing parameter, which controls the trade-off between the smoothness and goodness-of-fit of the fitted model. Therefore to control the influence of Z, the penalised likelihood is maximized:

$$[y^T(X\beta + Zu) - 1^T c(X\beta + Zu)] - \frac{1}{2}\lambda^2 \begin{bmatrix} \beta \\ u \end{bmatrix}^T D \begin{bmatrix} \beta \\ u \end{bmatrix}, \tag{9}$$

where y is the response vector, D is the positive semi-definite penalty matrix (Wahba 1978; Green and Silverman 1994), 1 the unit vector and c is determined by the link function used in the GLM. The first term in (9) measures the goodness-of-fit while the second term is the roughness penalty. λ is the smoothing parameter for which large values produce smoother curves while smaller values produce more wiggly curves.

The choice of base function, selection of knots, and the way penalization is done is determined by the smoothness of the penalised spline. The base functions include polynomial, truncated polynomial, and B-spline function. However, the choice of the base does not change the fit though some bases are numerically stable and allow computation of fit with greater accuracy. The reason for selecting one base over another is ease of implementation and interpretability. The choice of knots is mostly done by taking equidistant over the range of the covariate space, or based on the quantiles of the covariate distribution (Ruppert et al., 2003) but user defined criteria can be used too. Penalization is done in a variety of ways, like penalizing for large finite differences of adjacent coefficients or for large curvatures. The trade off between smoothness and closely matching the data is governed by the smoothing parameter. The type of smoothing parameter selected is very important in the application of splines. A number of methods are available that can be used to select a smoothing parameter. In this analysis, unbiased risk estimation (UBRE) was used.

3.5 Estimating force of infection

The force of infection is one of the primary epidemiological parameters of infectious diseases. Therefore, under the assumption of lifelong immunity and that the disease is in a steady state,

the sero-prevalence and the force of infection can be estimated from sero-prevalence data (Grenfell and Anderson, 1985). The prevalence of a disease in a statistical population is given as the ratio of sero-positives at a given exposure time to the total number of individuals in the population. Force of infection is the risk per time unit for an uninfected (that is, sero-negative) drug user to become infected. Let $\pi(d)$ be the prevalence of a disease (HCV or HBV) at exposure time (d_i). Then the force of infection is given by

$$\lambda(d) = \frac{\pi'(d)}{1 - \pi(d)}, \tag{10}$$

where $\pi'(d)$ is the derivative of the prevalence with respect to duration (exposure time). $\pi(d)$ is the cumulative distribution function of exposure time at infection.

4 Results

4.1 Statistical results

In this study, the results presented are based on cubic regression splines since they have a direct parameter interpretation and the basis does not require any re-scaling of the predictor variables before it can be used to construct a GAM (Wood 2006). Therefore, a model with cubic splines was fitted with both logit and complementary log-log (clog-log) links to evaluate the link that fits the data better. The best link was selected using Akaike Information Criterion (AIC), where the link with the lowest AIC was considered best. Therefore, HBV and HCV were considered as the response, the models with the logit link had lower AIC values 543.99, 544.01 and 984.062, 985.464 respectively. Therefore, the results presented in this report are based on cubic regression spline bases with a logit link and the smoothing parameter is selected automatically using unbiased risk estimation (UBRE).

4.2 Hepatitis B model

The methods described in Section 3 were applied to the data with hepatitis B as the response under the different assumptions. Therefore, using the model selection criteria described above the following predictors were in line with assumptions of Model 1: ever injected drugs, current injecting, being homosexual, history of being in prison, sharing injections, sharing sniffing

materials, education levels, and belonging to a particular treatment center. In contrast, HIV status and being vaccinated for HBV can best be modelled with the assumptions of Model 2.

Lastly, gender was modelled under the assumptions of model 3. The following variables are not important risk factor for hepatitis B, sharing injection, sharing sniffing material and being vaccinated for hepatitis B. Also, it is worth mentioning that the interaction between HIV status and exposure time on drug was significant.

Table 1 shows log odds estimates, standard errors and p-values. Therefore, taking the exponent of the beta, the odds of having Hepatitis B infection for people who have ever injected drugs is 3 times that of those who have never injected drugs (P=0.002). Also the odds of having hepatitis B for individuals in treatment centre 2 are at least 20% less than those who are in treatment center 1.

The interaction between HIV status and exposure time shows that the longer an individual is exposed to drugs the lesser are the chances of becoming infected with hepatitis B, that is 69% compared to individuals who use drugs for a short period of time.

Table 1: Parameter Estimates and Standard Errors for the final models for Hepatitis B

Parameter	**Model Type**	**Estimate(SE)**	**P-value**
Ever injected drugs (ref : Never injected drugs)	Model 1	1.107(0.358)	0.002*
Current injecting drugs (ref : Not current injecting drugs)	Model 1	0.592(0.240)	0.013*
Homosexual (ref : Non homosexual)	Model 1	1.075(0.627)	0.0087*
Been in prison (ref : Never been in prison)	Model 1	1.277(0.362)	0.000*
Share injections (ref : No sharing injection)	Model 1	0.282(0.309)	0.362
Share sniffing materials (ref : Don't share sniffing materials)	Model 1	-0.335(0.297)	0.258
HIV status (ref : HIV negative)	Model 2	6.429(2.057)	0.002*
Exposure time		0.064(0.035)	0.068
HIV status * Exposure time		-0.369(0.134)	0.006*
Education level (ref : low education level)	Model 1	-0.681(0.281)	0.015*
Vaccinated HBV (ref : No vaccination HBV)	Model 2	1.777(1.468)	0.226
Exposure time		0.144(0.068)	0.033*
Exposure time* Vaccinated HBV		-0.140(0.086)	0.103
Intercept (gender)	Model 3	-1.823(0.409)	8.37E-06*
Treatment centers			
CIC[2](ref :AC+MSOC)[1]	Model 1	-1.504(0.615)	0.014*
WGC+CGG+TG[3]		-0.411(0.350)	0.241
PAAZH+PH[4]		-0.473(0.351)	0.178

* Significant at 5% level

SE Standard error

[1] AC=Non-Residential day care centre,MSOC=Medical Social centres for drug users

[2] Crisis Intervention Centre

[3] CGG+WGC=Centres for Mental Health Care,TG=Therapeutic Community

[4] PAAZ=Psychiatric Unit within General Hospital,PH=Psychiatric Hospital

Figures 2 - 4 show how the prevalence for the different predictors changes over time of exposure for the different models. In general, the sero-prevalence of HBV by exposure time for all predictors was low where the majority of the predictors had sero-prevalence of 40% on average, except for homosexuals, sharing injections and HIV positive IDUs. A higher prevalence of HBV was observed in the following IDUs: those individuals who have ever injected drugs before, current injecting drugs, being homosexual, being in prisons, not being vaccinated for HBV, sharing injections, sharing sniffing materials, being HIV positive, being male, and being in a particular treatment center. This implies that IDUs in those groups are at a high risk

of contracting the infection the longer they are exposed to drugs. The plots indicate that the sero-prevalence of HBV is almost zero on average for all predictors between exposure times 0 to 15, then it suddenly increases rapidly between exposure times of around 15 and 20, whereafter it drops steadily downwards, as shown in Figure 2 - 4.

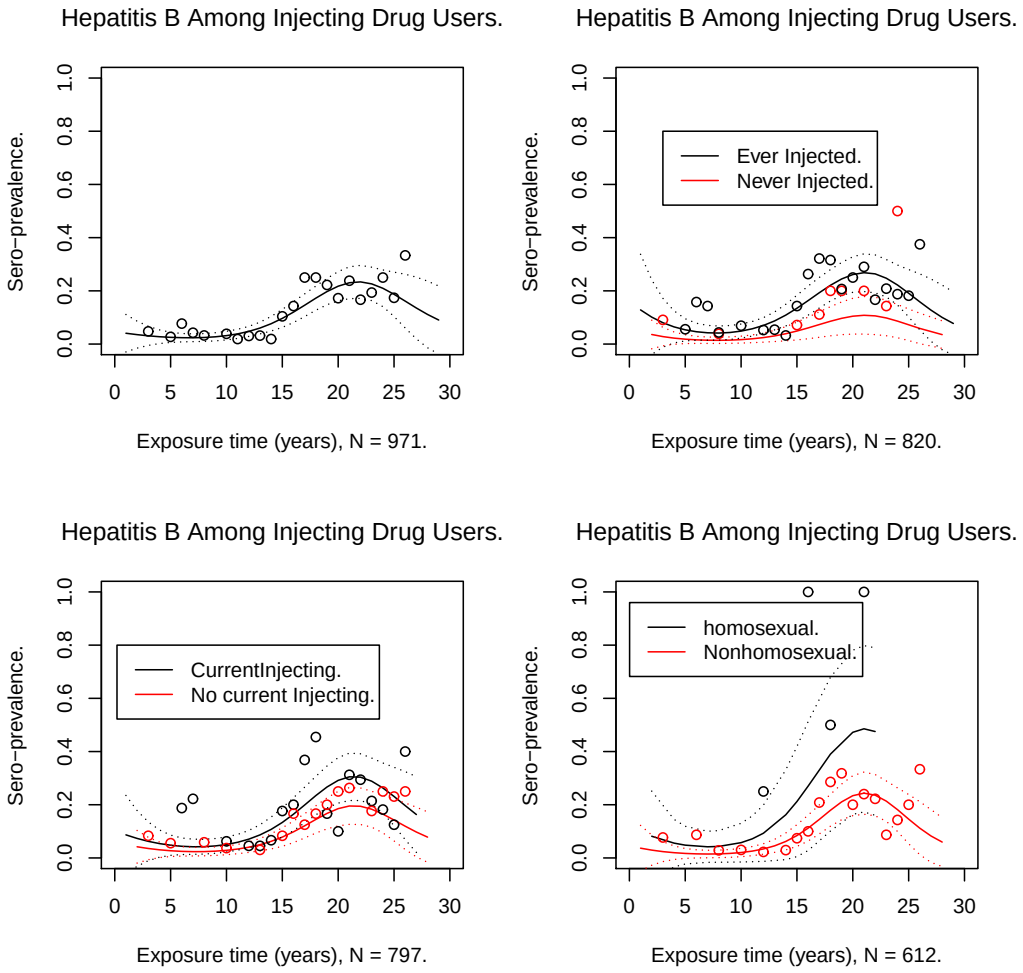


Figure 2: Fitted sero-prevalence (solid lines) for HBV infection. The dots are observed sero-prevalence and the dashed lines are confidence intervals.

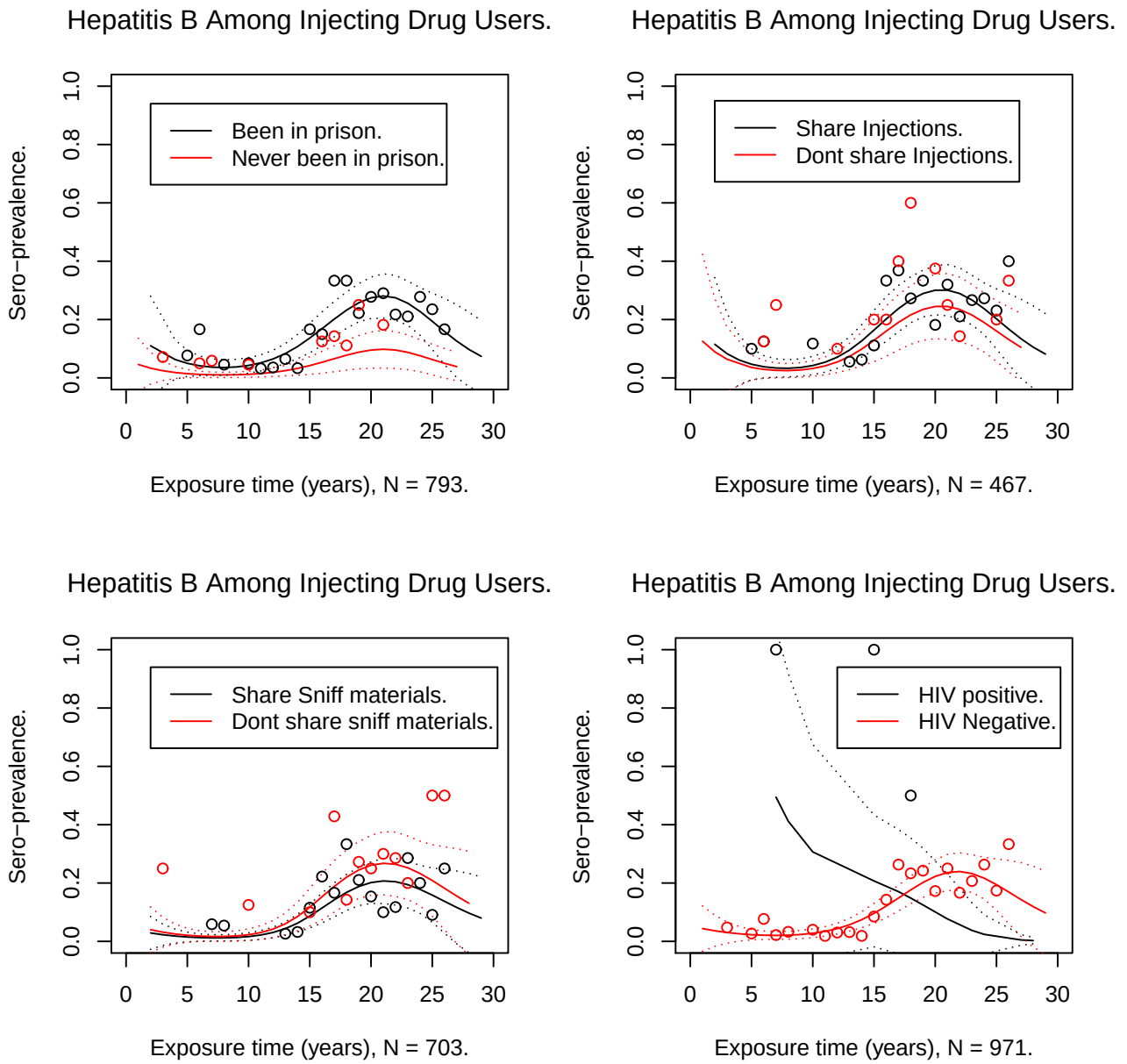


Figure 3: Fitted sero-prevalence (solid lines) for HBV infection. The dots are observed sero-prevalence and the dashed lines are confidence intervals.

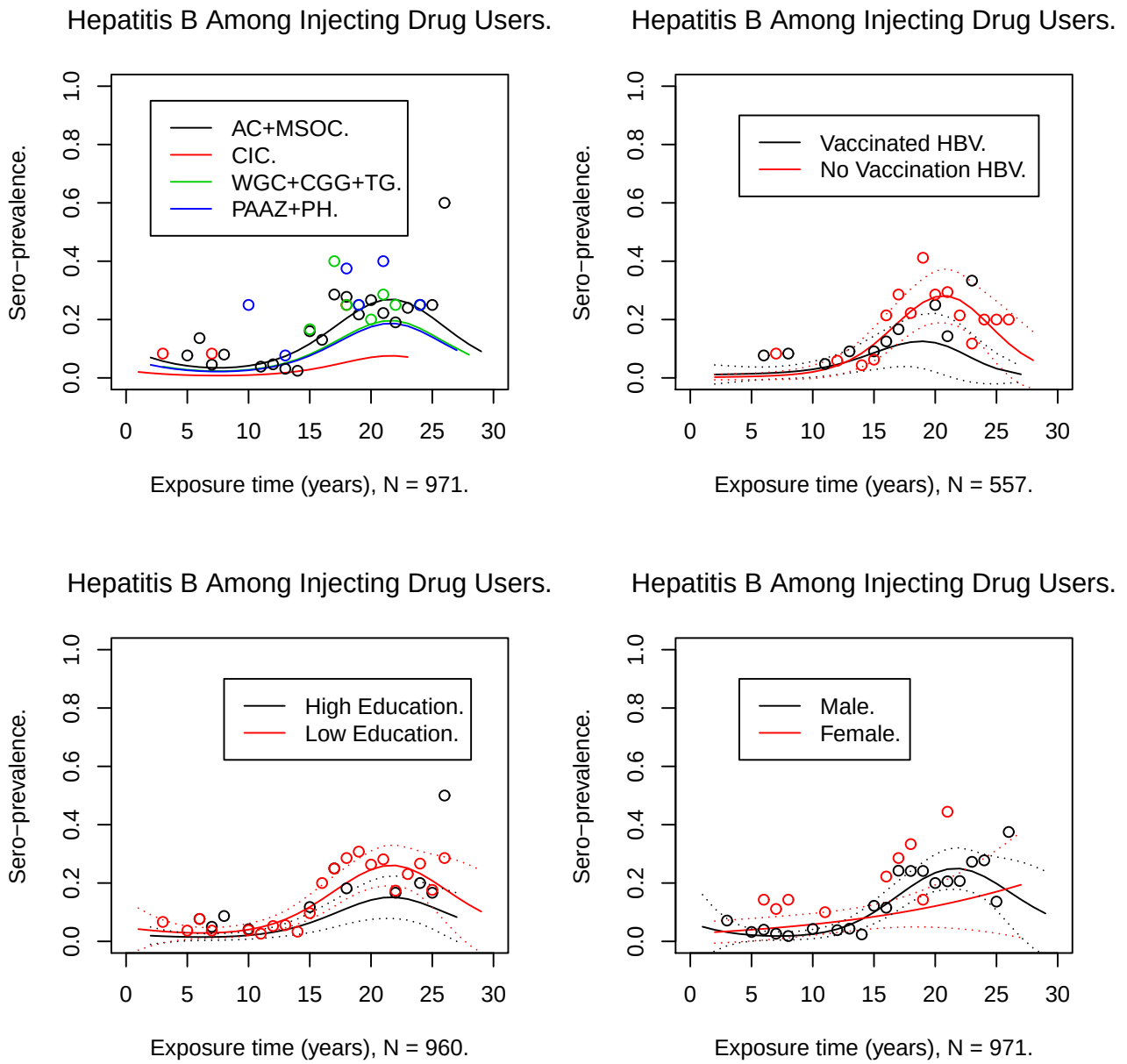


Figure 4: Fitted sero-prevalence (solid lines) for HBV infection. The dots are observed sero-prevalence and the dashed lines are confidence intervals.

4.3 Hepatitis C model

The hepatitis C Model was fitted under the same assumptions as mentioned in Section 2 above. Therefore using the model selection criteria described above, 'being homosexual' was found not to be an important risk factor for hepatitis C. The interaction between current injecting drugs and exposure time on drug was significant for hepatitis C.

Table 2: Parameter Estimates and Standard Errors for the final models for Hepatitis C

Parameter	**Model Type**	**Estimate(SE)**	**P-value**
Ever injected drugs (ref : Never injected drugs)	Model 1	2.871(0.327)	<2e-16*
Current injecting drugs (ref : Not current injecting drugs)	Model 2	2.79(0.557)	5.48E-07*
Exposure time		0.215(0.025)	< 2e-16*
Current injecting drugs*Exposure time		-0.068(0.033)	0.0405*
Homosexual(ref : Non homosexual)	Model 2	2.196(1.245)	0.078
Exposure time		0.173(0.018)	<2e-16*
Homosexual*Exposure time		-0.159(0.091)	0.081
Been in prison (ref : Never been in prison)	Model 1	0.782(0.190)	3.73E-05*
Share injections (ref : No sharing injection)	Model 1	0.982(0.226)	1.35E-05*
Intercept (gender)	Model 3	0.586(0.287)	0.042*
Share sniffing materials (ref : Don't share sniffing materials)	Model 1	-0.436(0.209)	0.037*
HIV status (ref : HIV negative)	Model 1	1.905(0.621)	0.002*
Education level (ref : low education level)	Model 1	-0.794(0.180)	9.87E-06*
Treatment centers			
CIC[2](ref:AC+MSOC)[1]	Model 1	-0.59(0.262)	0.024*
WGC+CGG+TG[3]		-0.812(0.250)	0.001*
PAAZH+PH[4]		-0.536(0.242)	0.026*

* Significant at 5% level

SE Standard error

[1] AC=Non-Residential day care centre,MSOC=Medical Social centres for drug users

[2] Crisis Intervention Centre

[3] CGG+WGC=Centres for Mental Health Care,TG=Therapeutic Community

[4] PAAZ=Psychiatric Unit within General Hospital,PH=Psychiatric Hospital

Table 2 above shows log odds estimates, standard errors and the p-values. From Table 2, taking the exponent of the beta for ever injected predictor, the odds of having hepatitis C infection for people who have ever injected drugs is almost 18 times higher than for those who have never injected drugs (P <0.001). While also the odds of having hepatitis C for individuals in treatment centre 2 are at least 55.4% less than for those who are in treatment center 1. The interaction between current injecting drugs and exposure time shows that the longer an individual is exposed to drugs the lesser are the chances of becoming infected with hepatitis C that is 93.4% compared to individuals who use drugs for a short period of time.

Furthermore, for HCV models the assumption that the difference in prevalence among the groups (say male and female) if it exists, does not depend only on the duration of the exposure on drugs (Model 1) but also on the group to which IDUs belongs. This assumption was satisfied by the following predictors: ever injected drugs, sharing sniffing materials, history of being in prison, education levels, sharing injections, HIV status and being in a particular treatment center. While current injecting and being homosexual can best be modelled with the assumptions of Model 2 and gender with assumptions of Model 3.

To get a better insight on how the prevalence changes over time of exposure the models fitted above are presented graphically below. Generally the sero-prevalence of drug users infected with HCV increases with longer exposure time on drugs for all the predictors. The HCV prevalence is higher among all potential predictors compared to HBV. A higher prevalence of HCV was observed in the following IDUs Individuals: those who have ever injected drugs before, current injecting drugs, being in prisons, sharing injections, not sharing sniffing materials, being HIV positive, low level of education, being male, and being in a particular treatment center. Thus, this implies that IDUs in those groups are at a high risk of contracting the infection the longer they are exposed to drugs. The plots show that the prevalence of HCV increases steadily for all potential predictors for longer exposure times. The median duration of the infection, that is the duration of exposure at which the sero-prevalence reaches 50% for the injecting drug user, can be estimated to be between 15 to 17 years of exposure for all predictors as shown in Figures 5 - 7.

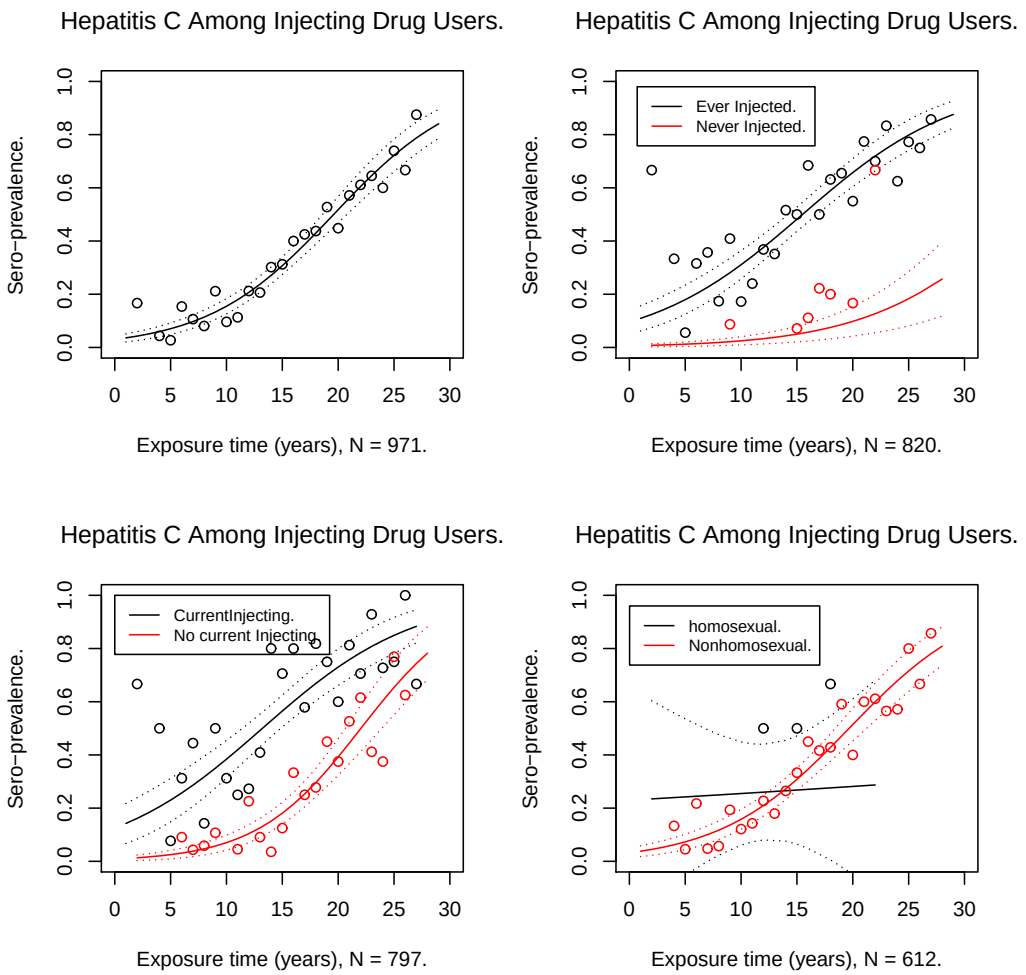


Figure 5: Fitted sero-prevalence (solid lines) for HCV infection. The dots are observed sero-prevalence and the dashed lines are confidence intervals.

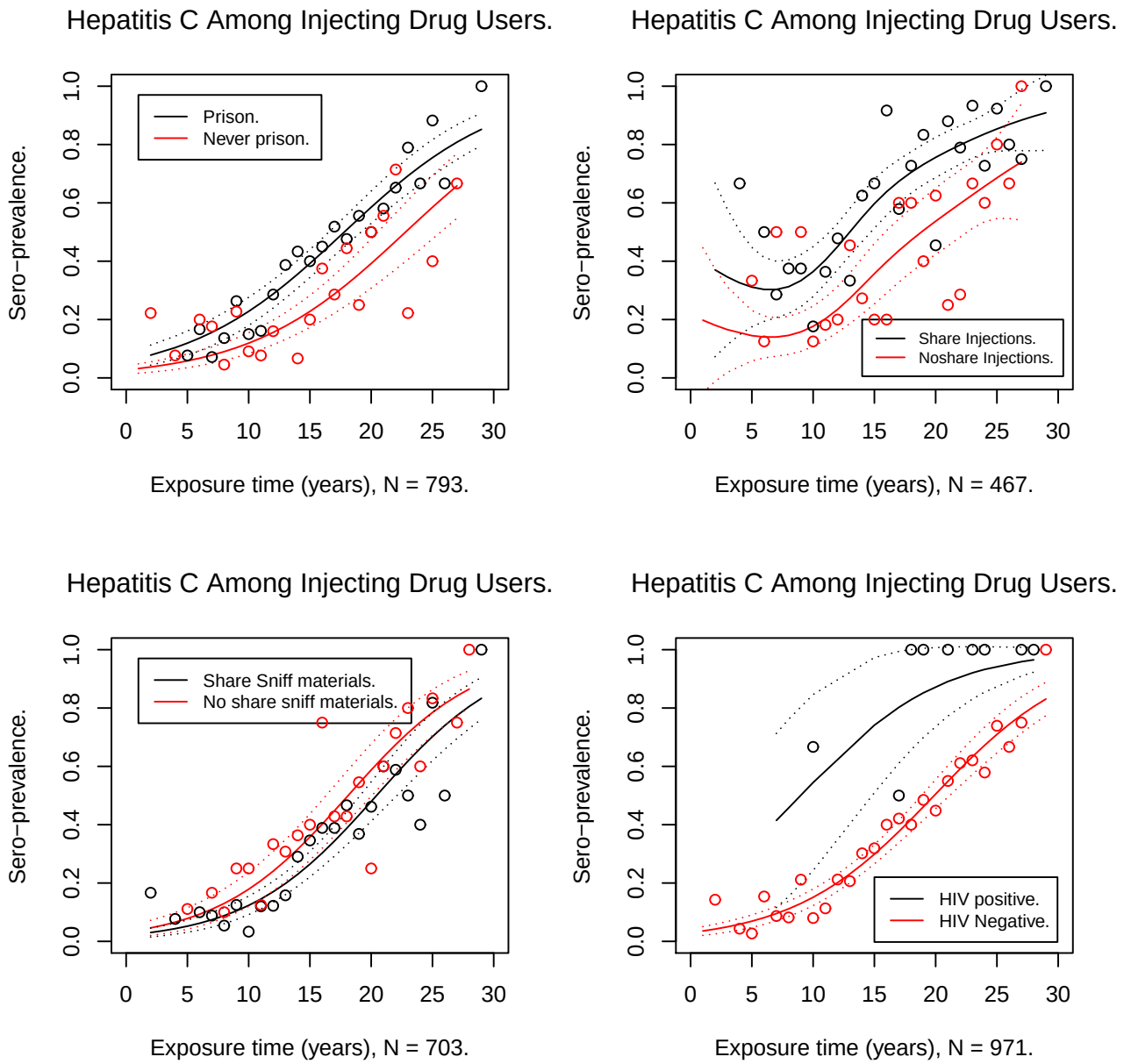
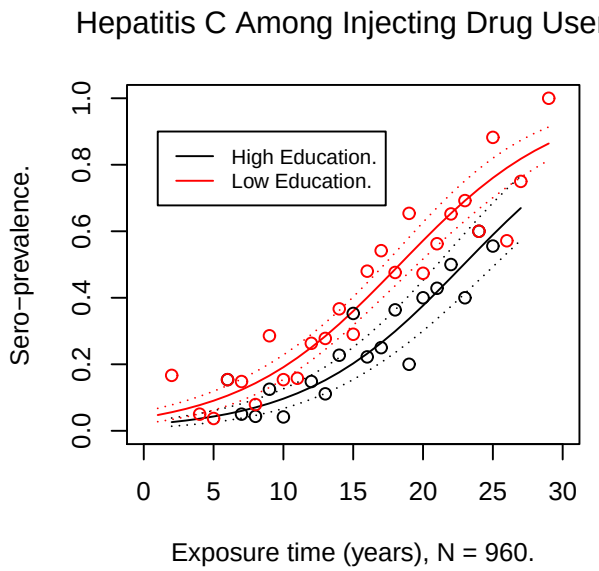
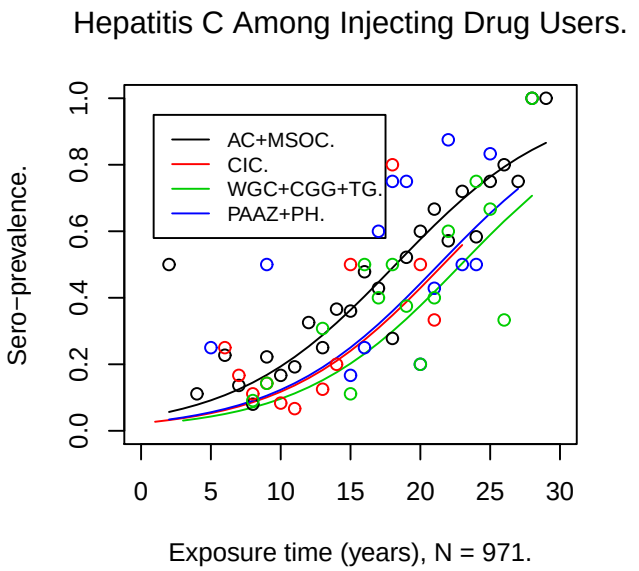


Figure 6: Fitted sero-prevalence (solid lines) for HCV infection. The dots are observed sero-prevalence and the dashed lines are confidence intervals.





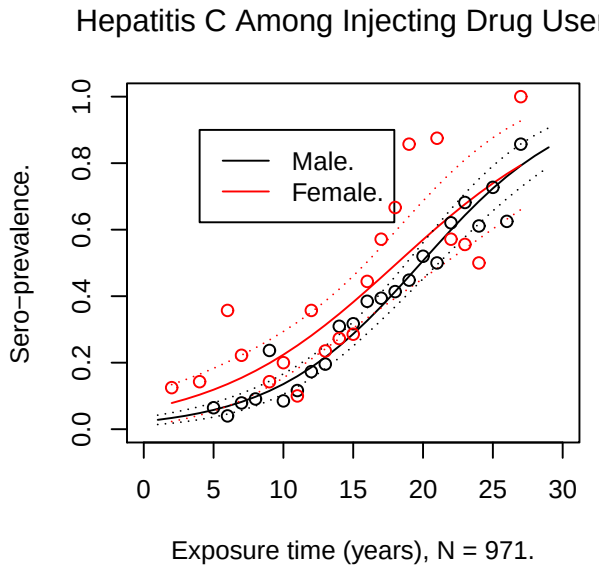


Figure 7: Fitted sero-prevalence (solid lines) for HCV infection.The dots are observed sero-prevalence and the dashed lines are confidence intervals.

5 Discussion

This was a study on HBV and HCV sero-prevalence in a sample of drug users in contact with treatment centres or in prisons in Belgium, 2004 - 2005. The objective was to estimate prevalence and also identify potential risk factors that are associated with HCV and HBV among injecting drug users. The analysis was carried out on IDUs in the age group of 15 to 40 years. The outcome variables studied are HBV and HCV status. HIV status was included as a risk factor because it had a low sample size. A unique feature of this study was the recruitment of IDU subjects which was done on voluntary basis and all centres willing to participate could participate. The present study shows a high prevalence of blood borne viral hepatitis among IDUs in Belgium. This concurs with epidemiological data which indicates that IDUs represent the largest risk group for HCV infection. Also, as the study conforms with literature stating that hepatitis is usually higher among IDUs than in other comparable non-IDU population strata.

Generalised additive models were applied to identify the possible risk factor associated with HBV - HCV from the cross-sectional data. There is a significant difference in the observed sero-prevalence among the different IDUs between the infections. The HCV prevalence infection is very high across the different predictors compared to HBV. For the HCV response, prevalence has a positive association with the exposure time given any risk factor. The reason may be that since HCV infection has no cure nor vaccination for preventative purpose, therefore sero-prevalence is high among injecting drug users. However HBV prevalence seems to be zero for a number of years of exposure (0 to 15 years) time to drug user for all the risk factors then it suddenly rises and then drops.

The HCV prevalence is high among Injecting drug users (IDUs), this may be viewed as an indicator of the sharing of injecting equipment, and consequently as an indicator of HIV risk. This is in line with the graphical display of those predictors. Furthermore, this study is consistent with other studies in identifying a high prevalence of HCV infection among IDUs, and a strong association with the duration of exposure time. The low sero-prevalence in the HBV

infection could be attributed to good needle exchange practices that came into place since the beginning of 2001 in Flanders. Also, since the study took place in 2004-2005, the drug users injecting for less than 15 years did not participate to the big injecting parties in the mid eighties. This could explain why the sero-prevalence for HBV infection is almost zero between 0-15years of exposure despite HBV having a vaccine. The confidence intervals for homosexuals Figure 2 bottom right side are quite wide for HBV infection, the reason is that not many homosexuals are exposed for such a long period of time so the function is not accurately estimated in this region.

Since force of infection is one of the primary epidemiological parameters of infectious diseases, the force of infection for the different HCV infection models was estimated (see appendix figures 8 - 10). From the figures, the force of infection clearly depends on the duration of drug use (exposure time). Also, it is seen that the longer an individual keeps using drugs, also the risk of contracting an infection increases as well. Furthermore, drug users in different categories have different risks. Therefore, no matter the category the drug user belongs too, the risk of infection remains throughout the period of drug use. The force of infection for the different models of HBV was estimated but resulted into negative force of infection at low and higher exposure time. This can be related to sero-prevalence being non-monotone over the duration of exposure.

In conclusion, to develop appropriate prevention strategies, it is important to identify risk factors associated with HCV and HBV infection among IDUs. The following risk factors were not significant to HBV infection: share injections, share sniffing materials, being vaccinated for HBV. Furthermore, being homosexual was not an important risk factor for contracting any of the infections. The interaction between HBV with HIV infection was significant this could probably be because of the similar transmission modalities.

6 References

Alter, M. J. (2006). Epidemiology of viral hepatitis and HIV co-infection. *Journal of Hepatology*, **44**, 6-9.

Anon. (1997). Hepatitis C global prevalence. *The Weekly Epidemiological Record*, **72**, 365-368.

Baker, B. (2002). *Natural History of Hepatitis C*. Bethesda: National Institutes of Health.

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

Crofts, N., Hopper, J. L., Scott, D., Breschkin, A. M., Milner, R. and Locarnini, S. A. (1993). Hepatitis C virus infection among a cohort of victorian injecting drug users. *The Medical Journal Australia*, **159**, 237-41.

Del Fava, E., Shkedy, Z., Hens, N., Aerts, M., Barbara, S., Laura, C., Fernando, V., Wiessing, L. and Mirjam, K. (2011). Joint modeling of HCV and HIV co-infection among injecting drug users in Italy and Spain using individual cross-sectional data. *Statistical Communications in Infectious Diseases*, **3**, 1-3.

Desenclos, J. C. (2004). The challenge of hepatitis C in Europe. *Eurosurveillance Monthly Release*, **8**, 99.

Devroey, D., Van Casteren, V. and Vranckx, R. (1994). *Incidence of Acute Viral Hepatitis in Belgium, Results of the General Practitioners Sentinel Survey 1991-1992*. Brussels: Scientific Institute of Public Health.

European Monitoring Centre for Drugs and Drug Addiction. (2003). *The State of the Drugs Problem in the European Union*. Lisbon: Publications Office of the European Union.

European Monitoring Centre for Drugs and Drug Addiction. (2010). *The State of the Drugs Problem in Europe*. Luxembourg: Publications Office of the European Union.

Flamm, S. L., Parker, R. A. and Chopra, S. (1998). Risk factors associated with chronic hepatitis C virus infection. *American Journal of Gastroenterology*, **93**, 597-600.

Hastie, T. and Tibshirani, R. (1990). *Generalised Additive Models.* London: Chapman and Hall.

Hens, N., Ziv, S., Aerts, M., Faes, C., Van Damme, P. and Beutels, P. (2010). *Modeling Infectious Disease Parameter Based on Serological and Social Contact Data.* Springer. [Unpublished].

Hou, J., Liu, Z. and Gu, F. (2005). Epidemiology and prevention of hepatitis B virus infection. *International Journal of Medical Sciences*, **2**, 50-57.

Hutse, V. and Quoilin, S. (2004). *Prevalentiestudie Van Hepatitis A, B en C in Vlaanderen via Speekselstoffen.* Brussels: Scientific Institute of Public Health.

Ingrid, B., Rachael, B., Michelle, S., Margaret, M. and John, K. (1994). Risk factors for hepatitis C virus infection among injecting drug users in Sydney. *Genitourin Medicine Journal*, **70**, 321-324.

Kretzschmar, M. and Wiessing, L. (2006). *Modelling the Transmission of Hepatitis C in Injection Drug Users.* Luxembourg: European Monitoring Center for Drugs and Drug Addiction.

Lauer, G. M. and Walker, B. D. (2001). Hepatitis C virus infection. *The New England Journal of Medicine*, **345**, 41-52.

Mahoney, F. J. (1999). Update on diagnosis, management and prevention of hepatitis B virus infection. *Clinical Microbiology Reviews*, **12**, 351-366.

Mathei, C., Shkedy, Z., Denis, B., Kabali, C., Aerts, M., Molenberghs, G., Van Damme, P. and Buntinx, F. (2006). Evidence for a substantial role of sharing of injecting paraohenalia other than syringes/needles to the spread of hepatitis C among injecting drug users. *Journal of Viral Hepatitis*, **13**, 560-570.

Mcmahon, J. M., Simm, M., Milano, D. and Clatts, M. (2004). *Detection of Hepatitis C Virus in the Nasal Secretions of an Intranasal Drug User.* Available From: http://www.ann-clinmicrob.com/content/3/1/6. [Accessed on 08 October, 2012].

Perz, J. F., Armstrong, G. L., Farrington, L. A., Hutin, Y. J. and Bell, B. P. (2006). The contributions of hepatitis B virus and hepatitis C virus infections to cirrhosis and primary

liver cancer worldwide. *Journal of Hepatology*, **45**, 529-538.

Proust, B., Dubois, F., Bacq, Y., Le Pogam, S., Rogez, S., Levillain, R. and Goudeau, A. (2000). Two successive hepatitis C infections in an intravenous drug user. *Journal of Clinical Microbiology*, **38**, 3125-3127.

Raes, V. and Lombaert, G. (2004). *Standard Tables De Sleutel Respectively for 1999, 2000, 2001, 2002, 2003, 2004.* Gent: Department Research and Quality Assurance.

Ruppert, D., Wand, M. P. and Carroll, R. J. (2003). *Semiparametric Regression.* Cambridge: Cambridge University Press.

Simon, N. W. (2006). *Generalized Additive Models an Introduction with R.* London: Chapman and Hall/CRC Press.

Sleiman, S. (2005). *Belgian National Report on Drugs.* Brussels: Scientific Institute of Public Health.

Sorrell, M. F., Belongia, E. A., Costa, J., Gareen, I. F., Grem, J. L., Inadomi, J. M., Kern, E. R., McHugh, J. A., Petersen, G. M., Rein, M. F., Strader, F. M., Strader, B. D. and Trotter, H. T. (2009). Management of hepatitis B. *Annals of Internal Medicine Journal*, **150**, 104-110.

Van, H., Haastrecht, A., Goudsmit, J., Coutinho, R. A. and De Wolf, F. (1990). Prevalence, incidence, and risk factors of hepatitis C virus infection among drug users in Amsterdam. *The Journal of Infectious Diseases*, **162**, 823-826.

Wahba, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *Journal of the Royal Statistical Society*, **40**, 364-372.

Wiessing, L., Hedrich, D., Taylor, C. and Griffiths, P. (2003). *Hepatitis C a Hidden Epidemic, Drugs in Focus.* Lisbon: European Monitoring Centre for Drugs and Drug Addiction.

World Health Organization. (2004). Hepatitis B vaccines. *The Weekly Epidemiological Record*, **79**, 255-263.

7 Appendix

Table 3: Risk factors for HBV and HCV

Variable	**Total**	**HCV$^+$(%)**	**HBV$^+$(%)**
Homosexual			
Yes	23	6(0.98)	4(0.65)
No	592	179(29.1)	52(8.46)
Prison			
No	314	57(7.13)	10(1.25)
Yes	485	208(26.03)	72(9.01)
Ever IDU			
No	277	11(1.33)	10(1.21)
Yes	549	272(32.93)	78(9.44)
Current IDU			
No	482	96(11.96)	38(4.73)
Yes	321	175(21.79)	47(5.85)
Share Injection			
No	146	52(11.06)	17(3.62)
Yes	324	196(41.7)	54(11.49)
Share sniff material			
No	184	67(9.48)	22(3.11)
Yes	523	129(18.25)	38(5.37)
Treatment Center			
AC+MSOC[1]	540	200(20.43)	67(6.84)
CIC[2]	173	24(2.45)	3(0.31)
WGC+CGG+TG[3]	134	29(2.96)	11(1.12)
PAAZ+PH[4]	132	35(3.58)	11(1.12)

[1] AC=Non-Residential day care centre,MSOC=Medical Social centres for drug users

[2] Crisis Intervention Centre

[3] CGG+WGC=Centres for Mental Health Care,TG=Therapeutic Community

[4] PAAZ=Psychiatric Unit within General Hospital,PH=Psychiatric Hospital

Table 4: Risk factors for HBV and HCV

Variable	**Total**	**HCV$^+$(%)**	**HBV$^+$(%)**
Gender			
Female	195	63(6.44)	15(1.53)
Male	784	225(22.98)	77(7.87)
HIV status			
Negative	960	273(27.89)	89(9.09)
Positive	19	15(1.53)	3(0.31)
Education Level			
Low	641	222(22.96)	73(7.55)
High	326	61(6.31)	18(1.86)
Vaccinated HBV			
No	326		41(7.31)
Yes	235		11(1.96)

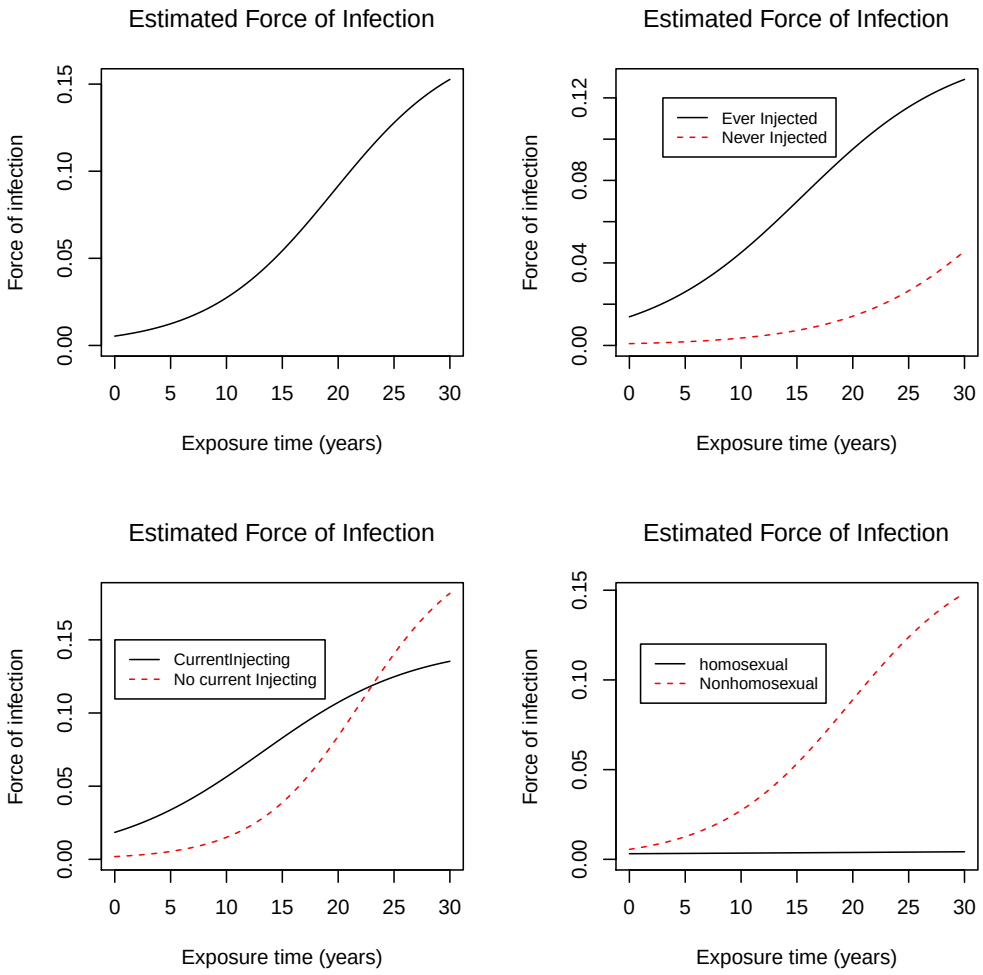


Figure 8: Force of Infection for Drugs Users infected with HCV infection.

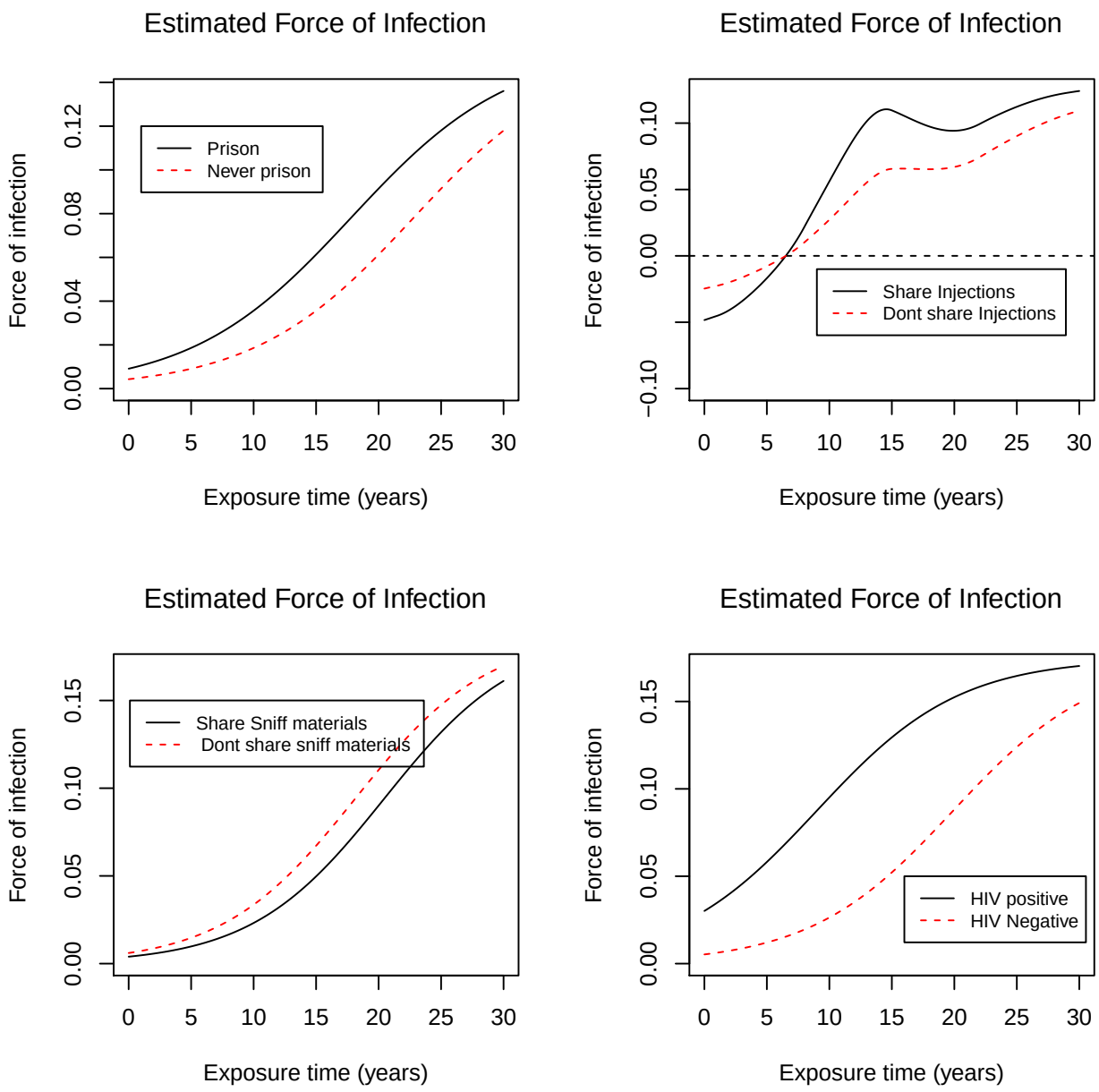


Figure 9: Force of Infection for Drugs Users infected with HCV infection.

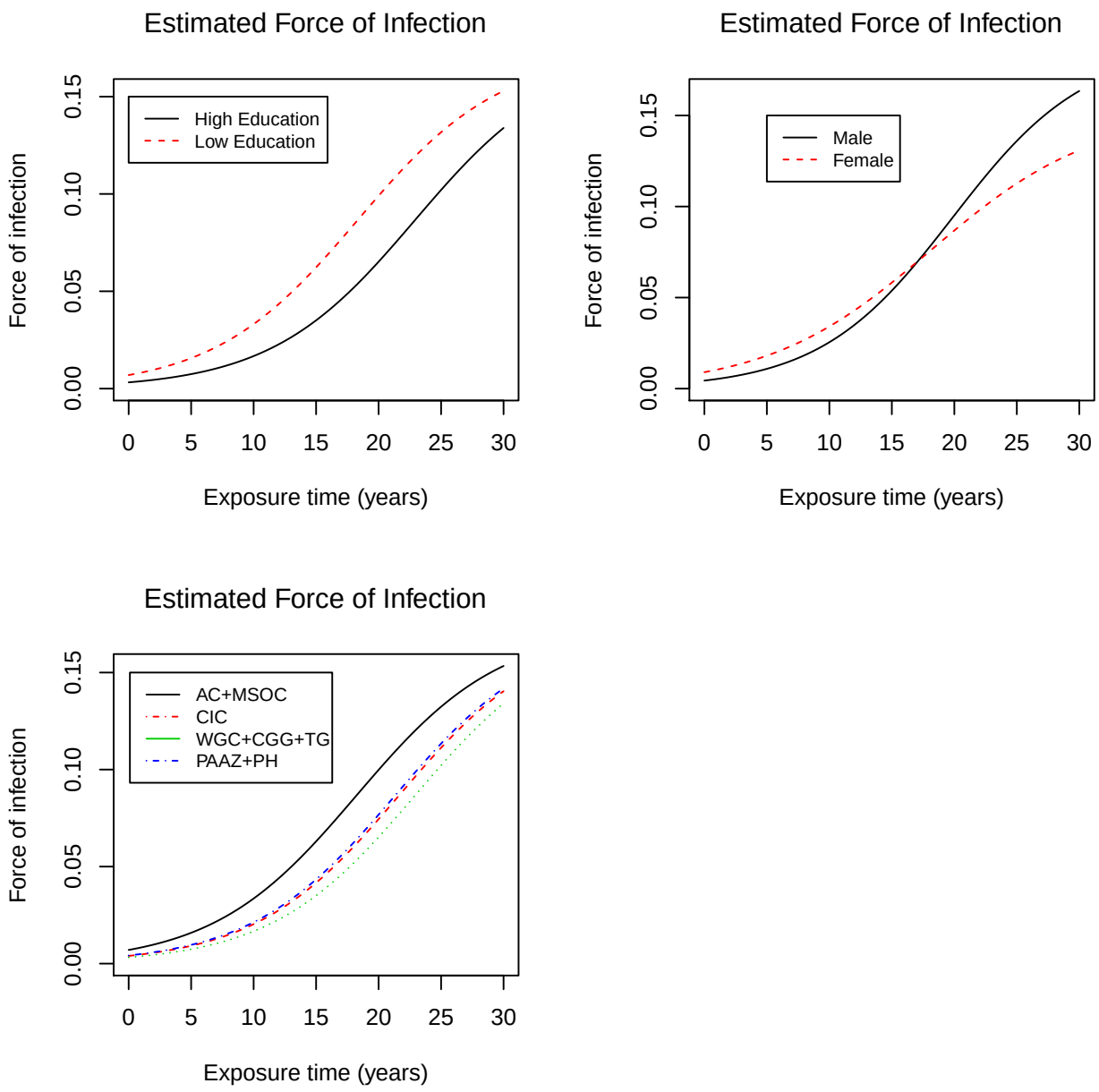


Figure 10: Force of Infection for Drugs Users infected with HCV infection.

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
Estimating Prevalence of Hepatitis B and C Among Injecting Drug Users and Identifying Risk Factors Associated with the Infections

Richting: **Master of Statistics-Biostatistics**
Jaar: **2013**

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,

Ntanda, Henry

Datum: **23/10/2012**