Joint Modeling of HCV and HIV Co-Infection among Injecting Drug Users in Italy and Spain Using Individual Cross-Sectional Data

Emanuele Del Fava, Hasselt University
Ziv Shkedy, Hasselt University
Niel Hens, Hasselt University and University of Antwerp
Marc Aerts, Hasselt University
Barbara Suligoi, Italian National Institute of Health
Laura Camoni, Italian National Institute of Health
Fernando Vallejo, Institute of Health Carlos III
Lucas Wiessing, European Monitoring Centre for Drugs and Drug Addiction
Mirjam Kretzschmar, Julius Centre and RIVM

Recommended Citation:
Del Fava, Emanuele; Shkedy, Ziv; Hens, Niel; Aerts, Marc; Suligoi, Barbara; Camoni, Laura; Vallejo, Fernando; Wiessing, Lucas; and Kretzschmar, Mirjam (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: Vol. 3: Iss. 1, Article 3.
DOI: 10.2202/1948-4690.1010
Joint Modeling of HCV and HIV Co-Infection among Injecting Drug Users in Italy and Spain Using Individual Cross-Sectional Data

Emanuele Del Fava, Ziv Shkedy, Niel Hens, Marc Aerts, Barbara Suligoi, Laura Camoni, Fernando Vallejo, Lucas Wiessing, and Mirjam Kretzschmar

Abstract

The aim of the analysis presented in this paper is to study co-infection with hepatitis C virus (HCV) and human immunodeficiency virus (HIV) in injecting drug users (IDUs) using a joint modeling approach that makes use of multivariate statistical methods for current status data.

Using marginal models, we estimate association measures between HCV and HIV infections at individual level, i.e., odds ratios and correlation coefficients, and we regress them against some risk factors, e.g., the length of the injecting career, the age at first injection, the ever sharing of syringes, and the frequency of current injecting. In addition, we fit random-effects models that take into account the individual heterogeneity in the acquisition of the infections. For our analysis, we use cross-sectional data from two independent serological surveys, one carried out in Italy (IT) in 2005 on 856 subjects, and the other in three Spanish (ES) cities, between 2001 and 2003, on 589 subjects.

We found that the infections are positively associated within individuals, e.g., \( \text{OR}_{\text{IT}} = 2.56 \) with 95% confidence interval (CI) (1.43, 6.68) and \( \text{OR}_{\text{ES}} = 2.42 \), with 95% CI (1.41, 4.30). We found that the odds ratio and the correlation between HCV and HIV infections increase positively with the length of the injecting career. Moreover, they are found to be significantly positive in case IDUs have never shared syringes or report low injecting frequencies. The variance of the individual random effects is positive, e.g., \( \sigma^2 = 0.34 \) (0.14, 0.62), indicating that there is significant individual heterogeneity in the acquisition of the infections.

Our results show that a significant association between HCV and HIV infections within IDUs is related to significant individual heterogeneity in the acquisition of the infections. Indeed, the association between these infections in IDUs who report ever sharing syringes is not significant, which can be explained by a higher homogeneity in their behaviors and, therefore, in their acquisition of the infections.

**KEYWORDS:** current status data, HCV and HIV co-infection, odds ratio, individual heterogeneity, marginal models, mixed-effects models
Author Notes: Emanuele Del Fava, Interuniversity Institute for Biostatistics and statistical Bioinformatics, Hasselt University, Belgium. Ziv Shkedy, Interuniversity Institute for Biostatistics and statistical Bioinformatics, Hasselt University, Belgium. Niel Hens, Interuniversity Institute for Biostatistics and statistical Bioinformatics, Hasselt University, Belgium. Marc Aerts, Interuniversity Institute for Biostatistics and Statistical Bioinformatics, Hasselt University, Belgium. Barbara Suligoi, National AIDS Unit, Department of Infectious Diseases, National Health Institute, Rome, Italy. Laura Camoni, Department of Infectious Diseases, National Health Institute, Rome, Italy. Fernando Vallejo, AIDS and Risk Behaviours Unit, National Centre of Epidemiology, Institute of Health Carlos III, Madrid, Spain. Lucas Wiessing, European Monitoring Centre for Drugs and Drug Addiction, Lisbon, Portugal. Mirjam Kretzschmar, Julius Centre, University Medical Centre Utrecht, and RIVM, Centre for Infectious Disease Control, Bilthoven, Netherlands. This study contributes to the work of the “European Study Group for Mathematical Modeling and Epidemiological Analysis of Drug-Related Infectious Diseases,” coordinated by EMCDDA and RIVM with funding from WHO/Europe and the government of the Netherlands. EDF thanks Prof. Piero Manfredi for the precious and insightful scientific discussion on the matter that they had together in Pisa; EDF also thanks C. Paddy Farrington and Steffen Unkel for a valuable discussion on the matter and for their insightful comments on an earlier version of the paper. NH gratefully acknowledges financial support from “SIMID,” a strategic basic research project funded by the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT), project number 060081. EDF, ZS, NH, and MA gratefully acknowledge the IAP research network number P6/03 of the Belgian Government (Belgian Science Policy). EDF, ZS, NH, and MA gratefully acknowledge support from the Belgian IUAP/PAI network “Statistical techniques and modeling for complex substantive questions with complex data.” Finally, we thank the editor and the three anonymous referees for their valuable comments that helped us to improve substantially the content and the presentation of the manuscript.
1 Introduction

Hepatitis C is an infectious disease affecting the liver, caused by the hepatitis C virus (HCV). The infection is often asymptomatic, but once established, chronic infection can progress to scarring of the liver (fibrosis) and advanced scarring (cirrhosis). In some cases, those with cirrhosis will develop liver failure or other complications of cirrhosis, including liver cancer. No vaccine against HCV is 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). It has been found that the sexual transmission of HCV is rare (Vandelli et al., 2004); however, in case of men having sex with men, it has been seen recently that HCV infection can occur in subjects already HIV-infected in absence of injecting drug use (van de Laar et al., 2011), thus HCV infection emerges as a sexual transmitted infection in this population. The exposure to infected blood in the context of injecting drug use is the predominant way of transmission in the developed countries characterized by low and very low prevalence (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 (Mathëi et al., 2002; Shepard et al., 2005; Mathëi et al., 2006). The estimated prevalence of HCV infection worldwide is 2%, representing 123 million people (Perz et al., 2006). Italy and Spain are among the Western European countries with the highest prevalence for HCV and HIV infection. 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. Mathëi et al. (2002) report similar figures for the prevalence of HCV infection in IDUs.

Acquired immunodeficiency syndrome (AIDS) is a disease of the human immune system caused by the human immunodeficiency virus (HIV). This syndrome progressively reduces the effectiveness of the immune system and leaves individuals susceptible to opportunistic infections and tumors. HIV is transmitted through direct contact of a mucous membrane or the bloodstream with a bodily fluid containing HIV, such as blood, semen, vaginal fluid, preseminal fluid, and breast milk. This transmission can involve anal, vaginal or oral sex, blood transfusion, contaminated hypodermic needles, exchange between mother and baby during pregnancy, childbirth, or breastfeeding, or other exposure to one of the above mentioned bodily fluids. According to the Joint United Nations Programme on HIV/AIDS (UNAIDS), the estimated number of people living with AIDS in 2008 is 33.4 million (0.8%), even though the number of newly infected is decreasing in recent years (UNAIDS, 2009). The prevalence of HIV infection in the general population in
Italy is between 0.1% and 0.5%, while the prevalence in Spain ranges between 0.5% and 1% (UNAIDS, 2008); in contrast, available data suggest a prevalence of HIV infection among IDUs of 11.7% and 34.5% (EMCDDA, 2010) in Italy and Spain, respectively.

Within the population of IDUs, HCV and HIV share the possibility of transmission by exposure with infected blood. HIV infection appears to accelerate HCV-related liver disease, while HCV infection does not appear to affect the rate of HIV disease progression. It is estimated that up to 70% of HIV-positive IDUs will also be HCV-positive due to their shared transmission route (Rockstroch and Spengler, 2004). Considering the general population, Alter (2006) reported that, among the HIV-infected persons, 4-5 million were also HCV-infected. If we compare the prevalence of HCV and HIV infections presented above, we notice that the former is much higher than the latter, both in the general population and in the group of IDUs. Moreover, it is estimated that the transmission risk from HIV-contaminated syringes is ten times less likely than if the syringes were contaminated with HCV, thus HIV requires more exposure to reach high prevalence (Hagan and Des Jarlais, 2000; Crofts et al., 2001; Alter, 2006).

In recent years, many epidemiological studies focussed on the co-infection with HCV and HIV in IDUs (Hagan and Des Jarlais, 2000; Alter, 2006). Kretzschmar and Wiessing (2008) remarked on the importance of using mathematical and statistical modeling in order to get a deeper insight in the epidemiological processes of this co-infection. A number of studies investigated the topic by analyzing data from cross-sectional surveys either in the general population (Sherman et al., 2002; Roca et al., 2003) or among IDUs from countries with different prevalence settings, usually concentrating on the effects of sexual and drug-related risk factors on the prevalence of both infections (Miller et al., 2004; Rhodes et al., 2005; Zoccratto et al., 2006; Bollepalli et al., 2007; Dumchev et al., 2009; Vickerman et al., 2009; Camoni et al., 2010; Rahimi-Movaghar et al., 2010; Vickerman et al., 2010). In this paper we focus on the association between HCV and HIV infections in the IDU population, using individual bivariate binomial data (the HCV and HIV infection statuses). Hence, we use statistical models that can take into account the clustered nature of these binomial data, and model the association either as a constant or as a function of known drug-related behavioral risk factors, such as the length of the injecting career, the age at first injection, the frequency of current injecting and the sharing of syringes (Mathéi et al., 2006). Our focus on the association between HCV and HIV infections is motivated by two reasons. First, we are interested in the association itself in order to understand which risk factors can enhance or diminish it. Second, we want to estimate the degree of individual heterogeneity in the acquisition of the infections. The idea is that each IDU carries a different risk of being infected, due, for instance, to a stronger or weaker genetical resistance to...
infections or due to residual confounding of unobserved behavioral risk factors that might work similarly for HIV and HCV infections or differentially between both infections. A way to obtain evidence of this individual heterogeneity is by analyzing data on co-infection with multiple viruses. If these infections are strongly associated within subjects, this provides evidence of heterogeneity in the way different subjects become infected, even though we do not identify the source of this heterogeneity (Coutinho et al., 1999; Farrington et al., 2001). To achieve this purpose, we consider two families of models. First, we use marginal models for binary data, namely, the alternating logistic regression (Carey et al., 1993), the bivariate Dale model (Dale, 1986), and the bivariate probit model (Ashford and Sowden, 1970; Morimune, 1979; Molenberghs and Verbeke, 2005), which allow for direct estimation of association measures between HCV and HIV infections (Hens et al., 2008). Second, we consider the family of mixed-effects models, i.e., the generalized linear mixed models (McCulloch and Searle, 2001; Molenberghs and Verbeke, 2005) and the Gamma frailty models (Farrington et al., 2001; Sutton et al., 2006 and 2008; Hens et al., 2009), which are conditional on random effects that capture the individual behavior in the acquisition of infections. With the latter models, we do not estimate directly the association, but we study the effect of the individual heterogeneity in becoming infected by testing the significance of the variance of the individual-specific random effects. All these statistical methods are applied to two serological cross-sectional samples of IDUs with exposure to several behavioral risk factors from Italy and Spain.

This paper is organized in the following way. Section 2 presents the data used in the analysis and descriptive statistics are used to assess the degree of the co-infection. In Section 3 we present a joint model for HCV and HIV infections using the length of the injecting career as exposure time. The different modeling approaches mentioned above are used to model the association between the infections and are discussed in Section 3.1 and Section 3.2. In Section 4, the proposed methods are applied to two cross-sectional samples from Italy and Spain. Finally, in Section 5 we discuss the results.

2 Data

The cross-sectional data used in this analysis consist of two seroprevalence samples of IDUs from Italy ($N = 856$) and Spain ($N = 589$).

The data from Italy come from a cross-sectional survey carried out in 2005 among IDUs and non-IDUs at public drug-treatment centers (DTCs) in 14 Italian regions (Camoni et al., 2010). The original survey included 1330 persons, of whom 1009 injected at least once in the last year. As concerns the serological status for
HCV and HIV infections, the persons were not tested directly during the survey, but rather the information was taken from their clinical records in the DTC, where subjects’ blood specimens were tested for antibodies against HCV and HIV. Since in this study we focus on the association between HCV and HIV infection, the following inclusion criteria were applied: (1) all subjects have serological results for both infections, (2) all subjects have information about their length of injecting career, (i.e., the age at first injection should be reported). Using these inclusion criteria, we obtained a final dataset of 856 IDUs. Note that in this dataset information about behavioral risk factors in Italy is available for the length of the injecting career and for the age at first injection.

The data from Spain were collected in a cohort study on 961 subjects, the "Itinerere Project", which was carried out between April 2001 and December 2003 in the metropolitan areas of Madrid, Barcelona, and Seville, among both street-recruited current injecting and non-injecting heroin users. Subjects from DTCs were excluded (Barrio et al., 2007). All the subjects in the sample were administered a questionnaire for the collection of socio-demographic, sexual behavior and drug use data; moreover, a blood specimen was obtained and tested for antibodies against HBV, HCV, and HIV. The application of the inclusion criteria mentioned above for Italy lead to a total sample of 589 IDUs with complete information about their drug-related behavioral risk factors: the length of the injecting career, the age at first injection, the sharing of syringes, and the frequency of current injecting. The latter risk factor consists of four categories, namely, "every day", "1–6 days per week", "less than weekly", and "never"; the last category refers to those drug users who have not injected in the last month. For the sharing of syringes we present the analysis on a "ever/never" basis. This implies that we assume time-homogeneity for this behavioral risk factor, which might be not the case. For this reason, in the supplementary material file (Section 5.2), we also present the results for sharing syringes based on the "current status" information (last month before the interview) for a comparison of the results.

The bar plots in Figure 1 present the four observed joint prevalences of HCV and HIV infections depending on the length of the injecting career, revealing a clear pattern of association between the prevalence and the duration of injecting, as already reported in the literature (Pallás et al., 1999; Hagan and Des Jarlais, 2000; Mathëi et al., 2006): the probability of either having a single infection or being co-infected with both viruses increases quickly with the length of injecting career, most of all in the first 5 years, while the probability of still being susceptible decreases. We also notice that the prevalence of HCV infection is far higher than the prevalence of HIV infection and almost all of those who are HIV-infected are HCV-infected as well (Rhodes et al., 2005).
Figure 1: Bar plots with the observed joint probabilities for HCV and HIV infections among IDUs in Italy and Spain depending on the length of the injecting career.
Table 1 shows descriptive statistics for the prevalence of the two infections, for their co-infection and for the risk factors as well as association measures between HCV and HIV infections. The proportion of males is 0.73 in Italy and 0.74 in Spain. As regards the prevalence estimates, they are fairly comparable to the figures in the literature, considering the high variability of the estimates (Aceijas and Rhodes, 2007; Mathers et al., 2008; EMCDDA, 2010): we notice that, although the median length of the injecting career in Italy is higher than in Spain (14 versus 6 years, respectively), which may be due to the exclusion in the Spanish study of IDUs in drug treatment, the overall prevalence of HIV infection in Spain is higher (26%) than in Italy (10%), indicating higher incidence in Spain and/or higher mortality in Italy. The estimated odds ratios for co-infection are equal to 5.5 with 95% confidence interval (CI) 2.2–17.5 for Italy and to 4.2 (95% CI: 2.4–8) for Spain. Both tetrachoric correlations (Molenberghs and Verbeke, 2005) are equal to 0.4 and highly significant. Similarly, the $\chi^2$-test statistics are found to be significant for both countries, indicating that there is significant association between HCV and HIV infections.

<table>
<thead>
<tr>
<th></th>
<th>Italy</th>
<th>Spain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prev. HCV</td>
<td>77% (CI: 74%–80%)</td>
<td>74% (CI: 70%–77%)</td>
</tr>
<tr>
<td>Prev. HIV</td>
<td>10% (CI: 8%–12%)</td>
<td>26% (CI: 22%–30%)</td>
</tr>
<tr>
<td>Prev. co-inf. HCV–HIV</td>
<td>10% (CI: 8%–12%)</td>
<td>23% (CI: 20%–27%)</td>
</tr>
<tr>
<td>Median age first inj. ($Q_{0.25}, Q_{0.75}$)</td>
<td>20 (18,24)</td>
<td>19 (17,22)</td>
</tr>
<tr>
<td>Median age at interview ($Q_{0.25}, Q_{0.75}$)</td>
<td>35 (30,41)</td>
<td>27 (24,29)</td>
</tr>
<tr>
<td>Median leng. inj. car. ($Q_{0.25}, Q_{0.75}$)</td>
<td>14 (7,19)</td>
<td>6 (3,10)</td>
</tr>
<tr>
<td>Ever shared syringes</td>
<td><em>Yes</em></td>
<td>-</td>
</tr>
<tr>
<td></td>
<td><em>No</em></td>
<td>-</td>
</tr>
<tr>
<td>Frequency</td>
<td><em>Every day</em></td>
<td>-</td>
</tr>
<tr>
<td>current</td>
<td><em>1-6 days/week</em></td>
<td>-</td>
</tr>
<tr>
<td>injecting</td>
<td><em>Less than weekly</em></td>
<td>-</td>
</tr>
<tr>
<td></td>
<td><em>Never</em></td>
<td>-</td>
</tr>
<tr>
<td>Pearson’s $\chi^2$ for co-infection</td>
<td>16.3 ($P &lt; 0.0001$)</td>
<td>27.6 ($P &lt; 0.0001$)</td>
</tr>
<tr>
<td>Odds ratio for co-infection, $\psi$</td>
<td>5.5 (CI: 2.2–17.5)</td>
<td>4.2 (CI: 2.4–8)</td>
</tr>
<tr>
<td>Tetrachoric corr. for co-infection, $\rho$</td>
<td>0.4 ($P &lt; 0.0001$)</td>
<td>0.4 ($P &lt; 0.0001$)</td>
</tr>
</tbody>
</table>

Table 1: The prevalence of HCV infection, HIV infection, and their co-infection in the sample (with 95% CI). Descriptive statistics for the behavioral risk factors and association measures for the co-infection with HCV and HIV. For more information about the tetrachoric correlation, see Section 3 in the supplementary material file.

### 3 Statistical Methods

The data collected in the two surveys are bivariate. Each IDU is a cluster with two measurements corresponding to the serological status of HCV and HIV infection.
Let \( y_i = (y_{i1}, y_{i2}), i = 1, \ldots, n \), be a vector of indicator variables representing the serological status for HCV and HIV infections, respectively, for the \( i \)th subject. Then, let \( d_i \) be the length of the injecting career for the \( i \)th subject, calculated as the difference between the age at interview and the age at first injection.

In what follows we present different statistical models for clustered data, focussing on the estimation of the association between the two infections. For this purpose, we fitted both marginal and mixed effects models. In particular, for the marginal models we considered the alternating logistic regression model (Carey et al., 1993), the bivariate Dale model (Dale, 1986), and the bivariate probit model (Ashford and Sowden, 1970; Morimune, 1979; Molenberghs and Verbeke, 2005). The mixed effects models used are the generalized linear mixed model for binary data (McCulloch and Searle, 2001; Agresti, 2002; Molenberghs and Verbeke, 2005) and the shared Gamma frailty model (Farrington et al., 2001; Sutton et al., 2006 and 2008; Hens et al., 2009).

3.1 Modeling the Association Between HCV and HIV Infection Using Marginal Models

3.1.1 Alternating Logistic Regression (ALR) and Bivariate Dale Model (BDM)

Carey et al. (1993) argued that the objectives of a multivariate analysis of binary data should include (1) the description of the dependency of each binary response on some covariates and (2) the characterization of the degree of association between pairs of responses and the dependence of this association on covariates. In this sense, the first two marginal models we consider, the ALR (Carey et al., 1993) and the BDM (Dale, 1986), model the association between HCV and HIV infections using the odds ratio, the former using a quasi-likelihood approach, the latter using full likelihood. Let \( \psi_i \) be the pairwise odds ratio between responses \( y_{i1} \) and \( y_{i2} \) and let the log odds ratio be defined as

\[
\log(\psi_i) = \log \left[ \frac{P(y_{i1} = 1, y_{i2} = 1)P(y_{i1} = 0, y_{i2} = 0)}{P(y_{i1} = 1, y_{i2} = 0)P(y_{i1} = 0, y_{i2} = 1)} \right] = \log \left[ \frac{\pi_{11}\pi_{00}}{\pi_{10}\pi_{01}} \right].
\]

(1)

Here, \( P(y_{i1} = 1, y_{i2} = 1) = \pi_{11} \) is the joint probability to be infected with both viruses. In the simplest case, \( \log(\psi_i) = c \), i.e., the pairwise log odds ratio is constant. However, the marginal models considered allow to model the dependence of the association upon other behavioral risk factors for HCV and HIV infections. For example, let \( x_{i} \) be a binary behavioral risk factor, e.g., ever sharing syringes. The log odds ratio can be modeled as \( \log(\psi_i) = \alpha_0 + \alpha_1 x_{i} \). The case in which the null hypothesis \( H_0 : \alpha_1 = 0 \) cannot be rejected is consistent with the idea that the...
association between HCV and HIV infection is the same across the levels of \( x_i \). Furthermore, if \( x_i \) is continuous, e.g., the age at first injection, a more flexible model such as a smooth differentiable function \( h(\cdot) \) can be used to model the log odds ratio, \( \log(\psi_i) = h(x_i) \). Both the ALR and the BDA can be formulated in the following way:

\[
\begin{align*}
  g(P(y_{i1} = 1)) &= h_1(x_{i1}), \\
  g(P(y_{i2} = 1)) &= h_2(x_{i2}), \\
  \log(\psi_i) &= h_3(x_{i3}).
\end{align*}
\]  

(2)

Here, \( g(\cdot) \) is a link function, \( x_{i1}, x_{i2} \) and \( x_{i3} \) are vectors of covariates associated with the marginal prevalence of HCV and HIV infections and the odds ratio, respectively, for the \( i \)th subject. Note that \( x_{i1}, x_{i2} \) and \( x_{i3} \) may not be identical.

Since the prevalence of HCV and HIV infections depends on the length of the injecting career \( d_i \) as well as on other risk factors, we follow the modeling approach of Mathéi et al. (2006) and fit a parametric model for the prevalence:

\[ h_j(d_i) = \beta_{0j} + \beta_{1j} \log(d_i) + \gamma z_i \]

where the parameters \( \beta_{0j} \) and \( \beta_{1j} \) (\( j = 1, 2 \)) are infection-specific intercepts and slopes, respectively, \( z_i \) is a covariate representing a behavioral risk factor and \( \gamma \) is its coefficient.

For the log odds ratio, we regress it against each risk factor at a time, using the following linear model without intercept:

\[ h_3(z_i) = \alpha z_i, \]

where \( \alpha \) is the coefficient of \( z_i \), which can be one the following behavioral risk factors: length of the injecting career and age at first injection (continuous), ever sharing syringes (binary), and frequency of current injecting (categorical), the latter two available only for Spain.

Three standard link functions for binary data were considered: the logit link, \( \log(\pi/(1 - \pi)) \), which implies a log-logistic distribution for the time spent in the susceptible class, when the time covariate is expressed on a log scale; the probit link, \( \Phi^{-1}(\pi) \), which implies a log-normal distribution for the time spent in the susceptible class; the complementary log-log link, \( \log(-\log(1 - \pi)) \), which implies a Weibull distribution for the time spent in the susceptible class.

### 3.1.2 Bivariate Probit Model (BPM)

Another possibility to model the association between responses within the marginal models family is to use a BPM (Ashford and Sowden, 1970, Morimune, 1979, 2006).
Molenberghs and Verbeke, 2005). We assume that the current status of the infections \( y_i = (y_{i1}, y_{i2}) \) is related to two latent variables \( w_i = (w_{i1}, w_{i2}) \) that represent the unknown individual antibody levels of HCV and HIV infections. It is further assumed that the antibody levels of HCV and HIV infections follow a bivariate normal distribution given by

\[
\begin{pmatrix}
w_{i1} \\
w_{i2}
\end{pmatrix} \sim \text{MVN} \left( \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Sigma \right),
\]

where \( \Sigma \) is a 2x2 covariance matrix given by

\[
\Sigma = \begin{pmatrix}
\sigma_{w_1}^2 & \rho \sigma_{w_1} \sigma_{w_2} \\
\rho \sigma_{w_1} \sigma_{w_2} & \sigma_{w_2}^2
\end{pmatrix}.
\]

The serological status of each individual for the \( j \)th disease, \( j = 1, 2 \), is defined by

\[
y_{ij} = \begin{cases} 
1 & w_{ij} > \tau_j \quad \text{seropositive}, \\
0 & w_{ij} \leq \tau_j \quad \text{seronegative}.
\end{cases}
\]

Here, \( \tau_j \) is a infection-specific cut-off point used to classify individuals either as seropositive or as seronegative. The BPM allows to model the correlation \( \rho \) between the latent normal variables \( w_i \) underlying the binary data \( y_i \). The model can be formulated using a probit link function for each of the two marginal probabilities, while the correlation is fitted with the ”rhobit” link (Yee, 2010), which is commonly used for parameters that lie between \(-1\) and \(1\),

\[
\begin{align*}
\Phi^{-1}(P(y_{i1} = 1)) &= \beta_{01} + \beta_{11} \log(d_i) + \gamma z_i, \\
\Phi^{-1}(P(y_{i2} = 1)) &= \beta_{02} + \beta_{12} \log(d_i) + \gamma z_i, \\
\log \frac{1+\rho_i}{1-\rho_i} &= \alpha z_i.
\end{align*}
\]

Similarly to the odds ratio \( \psi \) in the ALR and BDM, \( \rho_i \) can be either constant or depending upon some behavioral risk factors \( z_i \).

### 3.2 Modeling the Association Between HCV and HIV Infections Using Mixed-Effects Models

The marginal models discussed in the previous sections were focussed on the estimation of odds ratio and correlation as association measures for co-infection with HCV and HIV. In this section, we discuss a second modeling approach whereby we concentrate on individual heterogeneity in the acquisition of infections. Two types of mixed-effects models are used: (1) a generalized linear mixed model (GLMM, McCulloch and Searle, 2001; Agresti, 2002; Molenberghs and Verbeke, 2005) and (2) a shared Gamma frailty model (Farrington et al., 2001; Sutton et al., 2006 and 2008; Hens et al., 2009).
3.2.1 Generalized Linear Mixed Models (GLMM)

A GLMM is an alternative modeling approach for multivariate data that captures individual heterogeneity by conditioning on subject-specific random effects in the model. For this reason, we refer to it as a conditional model. We assume that infections may be associated at individual level and that this association could be explained by individual random effects. The variance of these random effects accounts for the variability not explained by the covariates in the model, and indicates that there are differences among individuals, giving consequently evidence of heterogeneity in the transmission. For the analysis presented here, we used the so-called shared random intercept model. Let \( b_i \) be a subject-specific random intercept assumed to be normally distributed, \( b_i \sim N(0, \sigma_b^2) \). Conditional on \( b_i \), we model the prevalence of HCV and HIV infections in the following way:

\[
\begin{align*}
    g(P(y_{i1} = 1 | b_i)) &= \beta_{01} + \beta_{11} \log(d_i) + b_i, \\
    g(P(y_{i2} = 1 | b_i)) &= \beta_{02} + \beta_{12} \log(d_i) + b_i.
\end{align*}
\]

The parameter of primary interest in the GLMM specified in (6) is the variance \( \sigma_b^2 \) of the shared random intercepts \( b_i \). Whenever the null hypothesis \( H_0 : \sigma_b^2 = 0 \) is rejected, it means that there is significant heterogeneity among the subject-specific intercepts, and this means that infections are associated at individual level. If the null hypothesis cannot be rejected, then the data are consistent with the infections being acquired independently: IDUs could still vary in their risks for HCV and HIV infections, but not jointly. Note that this model implies that, for a given value of a covariate, there is no difference among individuals in their risk for HCV or HIV infections. In this case, the population-averaged prevalence models are more appropriate.

In order to compare the estimated population-averaged prevalence obtained from the GLMM with the prevalence obtained from the marginal models presented in Section 3.1, we need to marginalize the mixed-effects model. The marginalization consists in the integration of the random effects based on numerical integration techniques or numerical averaging. However, a more practical approach is to generate randomly a large number \( M \) of realized values for the random effects \( b_i \) from \( N(0, \sigma_b^2) \) (Verbeke and Molenberghs, 2000). Then, the marginalized prevalences of HCV and HIV infections, given an injecting career of length \( d_i \), the estimated coefficients \( \hat{\beta} = (\hat{\beta}_{0j}, \hat{\beta}_{1j}) \), \( j = 1, 2 \), and the realized values of \( \hat{b}_i \), are provided by the following formulae:

\[
\begin{align*}
    P(y_{i1} = 1) &= \frac{1}{M} \sum_{i=1}^{M} P(y_{i1} = 1 | d_i; \hat{\beta}, \hat{b}_i), \\
    P(y_{i2} = 1) &= \frac{1}{M} \sum_{i=1}^{M} P(y_{i2} = 1 | d_i; \hat{\beta}, \hat{b}_i).
\end{align*}
\]
3.2.2 Shared Gamma Frailty Models

A second conditional model, proposed by Farrington et al. (2001), assumes that every IDU is infected differently from the others, that is, the force of infection (FOI), \( \lambda(d_i, b_i) \), depends on subject-specific random effects \( b_i \), or "frailties," which represent to what extent IDUs carry a higher or lower risk of infection (Coutinho et al., 1999), and on the length of the injecting career \( d_i \). Given the assumption of proportional hazards, the FOI can be written as \( b_i \lambda_0(d_i) \), where \( \lambda_0 \) is the baseline hazard. Therefore, the susceptible proportion for the infection \( j \) is given by

\[
S_j(d_i | b_i) = \exp \left( -b_i \int_0^{d_i} \lambda_0(j)(t) dt \right), \quad j = 1, 2. \tag{8}
\]

Similar to the GLMM, the use of multiple sera allows the estimation of the distribution of these frailties, providing us with a measure of the individual heterogeneity in acquiring infections. We assume \( b_i \) to be shared between the two infections and to have Gamma distribution, \( \Gamma(\theta, 1/\theta) \), where the heterogeneity parameter \( \theta \) measures the association between the two infections. Using shared frailties, we assume perfect correlation among them at the level of the linear predictors and common variance. The frailties \( b_i \) have expected value \( E(b_i) = 1 \) and variance \( Var(b_i) = 1/\theta \): the smaller \( \theta \), the larger the heterogeneity, and the more different the way individuals acquire the infections. For the current version of the model, we follow Sutton et al. (2006 and 2008) and Hens et al. (2009). Let \( \pi_{00}(d_i) \) be the probability that individual \( i \) with a length of injecting career of \( d_i \) years has not been infected with either virus; let \( \pi_{10}(d_i) \) be the probability that individual \( i \) with a length of injecting career of \( d_i \) years has been infected with HCV, but not with HIV; similarly we define the joint probabilities \( \pi_{01}(d_i) \) and \( \pi_{11}(d_i) \). The model for the probability of being still susceptible to both diseases, derived from (8) by applying the Laplace transform (Sutton et al., 2006), is given by

\[
\pi_{00}(d_i) = \left[ S_1^{-1/\theta} + S_2^{-1/\theta} - 1 \right]^{-\theta}. \tag{9}
\]

Reparameterizing the joint probability in terms of the cumulative FOI, \( \Lambda_j(d_i) = \int_0^{d_i} \lambda_0(j)(t) dt \), we obtain the following set of equations for the four joint probabilities:

\[
\begin{align*}
\pi_{00}(d_i) &= \left[ \exp \left( \frac{\Lambda_1(d_i)}{\theta} \right) + \exp \left( \frac{\Lambda_2(d_i)}{\theta} \right) - 1 \right]^{-\theta}, \\
\pi_{10}(d_i) &= \exp \left( \frac{\Lambda_2(d_i)}{\theta} \right) - \pi_{00}(d_i), \\
\pi_{01}(d_i) &= \exp \left( \frac{\Lambda_1(d_i)}{\theta} \right) - \pi_{00}(d_i), \\
\pi_{11}(d_i) &= 1 - \pi_{10}(d_i) - \pi_{01}(d_i) - \pi_{00}(d_i) .
\end{align*}
\tag{10}
\]
To solve these equations, it is necessary to assume a model for the FOI, for instance, a log-logistic, a log-normal, or a Weibull model. The unknown parameters can be estimated by maximizing the log-likelihood of the observations,

\[ L = \sum_d \{n_{00d} \log[\pi_{00}(d)] + n_{10d} \log[\pi_{10}(d)] + n_{01d} \log[\pi_{01}(d)] + n_{11d} \log[\pi_{11}(d)] \}. \]

The likelihood ratio test (LRT) can be used to check whether the frailty \( b_i \) can be included in the model or not. Similarly to GLMMs, in order to compare the prevalence from the shared Gamma frailty model with the prevalences from the previous marginal models, we can compute the marginal prevalence of the infections using the estimated joint probabilities: for instance, the marginal prevalence of HCV infection, \( P(y_{i1} = 1|d_i) \), is given by the sum of the probability of being only HCV-infected and the probability of being co-infected, \( \pi_{10}(d_i) + \pi_{11}(d_i) \).

### 4 Results

#### 4.1 Constant Measures for Association

In this section, we present the results obtained from the models wherein the association measures were assumed to be constant. The initial mean structure of the models included infection-specific intercepts and coefficients for the logarithm of the length of injecting career, \( \log(d_i) \), without any other risk factors. Model selection was performed with the Akaike’s information criterion (AIC) and the Bayesian information criterion (BIC), except for the ALR, for which the Quasi-likelihood information criterion (QIC, Pan, 2001) was used. Based on the information criteria, the best mean structure and the best link function were selected. Regarding models’ computation, the ALR was fitted in SAS with the procedure GENMOD, while the BDM and the BPM were fitted in R with package \texttt{VGAM} (Yee, 2010); the GLMM was fitted in SAS with procedure GLIMMIX, whereas the shared Gamma frailty model was fitted in R.

Table 2 presents the estimates of the association measures for the models with the smallest information criteria. For all the models but the GLMM, the best mean structure remained the starting one, while for the GLMM the mean structure was reduced by constraining the coefficient of \( \log(d_i) \) to be shared between the infections. Besides, since the AIC and the BIC for the GLMMs with logit and probit link functions were very close to each other, we showed the estimates of the variance of the random effects for both models. For each prevalence model, results indicate that the association between the infections is significant, even though the confidence interval for the variance parameter of the shared Gamma frailty model in Italy
is quite close to zero. We notice that odds ratios obtained from the ALR and BDM are similar in both countries. For GLMM, the model with smallest AIC and BIC for Italy is the model with probit link, while for Spain it is the model with logit link. The variance of the random effects was found to be significant for both countries (P-values of LRT for the inclusion of random effects are equal to 0.014 and 0.0014 for Italy and Spain, respectively). In the same way, the heterogeneity parameters for the shared Gamma frailty model were found to be significant (P-values for the LRT for the inclusion of random effects are equal to 0.023 and 0.0023 for Italy and Spain, respectively). We notice that, due to scarcity of HIV infection cases in Italy, the 95% profile-likelihood CI for θ is very wide (1.07–176.43).

The results for Spain presented in this section were obtained for the complete cases analysis, which took into account only the individuals with complete information for all the behavioral risk factors (n = 589). For the sake of comparison, in Table 4 of the supplementary material file, we presented the results for Spain obtained for the available cases analysis, which considered all the individuals with complete information for the length of injecting career only (n = 619). Those results led to the same conclusion.

4.2 Testing for Common Variance for Subject-Specific Random Effects Between Countries

In this section we discuss an extension of the GLMM and the shared Gamma frailty model to the case of two populations. The parameter estimates for $\sigma^2_b$ reported in Table 2 suggest that the variance of random intercepts in the GLMM and in the shared Gamma frailty model may be equal for Italy (IT) and Spain (ES). Therefore, the two random-effects models were re-fitted for the two countries jointly and the null hypothesis of common variance of the random effects was tested.

For the GLMM, let $I_i$ be an indicator variable which takes value 1 if the IDU is from Italy and 0 otherwise. In order to test the null hypothesis, $H_0 : \sigma^2_{b,IT} = \sigma^2_{b,ES}$, for $j = 1, 2$, we formulated a joint linear predictor given by

$$h_j = \beta_{0jIT}I_i + \beta_{0jES}(1-I_i) + \beta_{1IT}I_i\log(d_i) + \beta_{1ES}(1-I_i)\log(d_i) + b_{IT}I_i + b_{ES}(1-I_i).$$

The parameters $\beta_{0jIT}, \beta_{0jES}, \beta_{1IT}$ and $\beta_{1ES}$ are infection and country-specific intercepts and slopes, respectively. Under the null hypothesis, the random effects $b_{IT}$ and $b_{ES}$ are assumed to follow a bivariate normal distribution with independent covariance matrix and common variance $\sigma^2_b$. The alternative hypothesis states that the variances of random effects are country-specific. Hence, the joint distribution of the random effects is given by
<table>
<thead>
<tr>
<th>Model</th>
<th>Association</th>
<th>Italy</th>
<th>Spain</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALR</td>
<td>$\psi$</td>
<td>2.58 (1.02, 6.53)</td>
<td>2.43 (1.35, 4.36)</td>
</tr>
<tr>
<td></td>
<td>link function</td>
<td>cloglog</td>
<td>cloglog</td>
</tr>
<tr>
<td>BDM</td>
<td>$\psi$</td>
<td>2.56 (1.43, 6.68)</td>
<td>2.42 (1.41, 4.30)</td>
</tr>
<tr>
<td></td>
<td>link function</td>
<td>cloglog</td>
<td>cloglog</td>
</tr>
<tr>
<td>BPM</td>
<td>$\rho$</td>
<td>0.23 (0.02, 0.44)</td>
<td>0.26 (0.12, 0.64)</td>
</tr>
<tr>
<td></td>
<td>link function</td>
<td>probit</td>
<td>probit</td>
</tr>
<tr>
<td>GLMM</td>
<td>$\sigma_b^2$</td>
<td>1.34 (0.23, 3.27)</td>
<td>1.15 (0.32, 2.44)</td>
</tr>
<tr>
<td></td>
<td>LRT</td>
<td>6.53 ($P = 0.0053$)</td>
<td>8.97 ($P = 0.0014$)</td>
</tr>
<tr>
<td></td>
<td>link function</td>
<td>logit</td>
<td>logit</td>
</tr>
<tr>
<td></td>
<td>$\sigma_b^2$</td>
<td>0.31 (0.026, 0.78)</td>
<td>0.36 (0.099, 0.74)</td>
</tr>
<tr>
<td></td>
<td>LRT</td>
<td>4.81 ($P = 0.014$)</td>
<td>8.72 ($P = 0.016$)</td>
</tr>
<tr>
<td></td>
<td>link function</td>
<td>probit</td>
<td>probit</td>
</tr>
<tr>
<td>Shared Gamma</td>
<td>$\sigma_b^2$</td>
<td>0.39 (0.01, 0.94)</td>
<td>0.37 (0.09, 0.73)</td>
</tr>
<tr>
<td>frailty model</td>
<td>$\theta$</td>
<td>2.59 (1.07, 176.43)</td>
<td>2.70 (1.38, 10.71)</td>
</tr>
<tr>
<td></td>
<td>LRT</td>
<td>3.99 ($P = 0.023$)</td>
<td>8.01 ($P = 0.0023$)</td>
</tr>
<tr>
<td></td>
<td>link function</td>
<td>cloglog</td>
<td>cloglog</td>
</tr>
</tbody>
</table>

Table 2: Estimates of the association measures and 95% CI. For each model and country, we reported the results for the model with the smallest AIC and BIC (except for the ALR, for which the QIC was used). The 95% CI for BDM and BPM are studentized bootstrap CI, based on $B = 999$ replicates; for ALR, we used asymptotic CI; for GLMM and shared Gamma frailty model, we used a profile-likelihood CI. For GLMM and shared Gamma frailty model, we report the result of LRT for the inclusion of a random intercept, as well. Information criteria are presented in Section 4 of the supplementary material file.

\[
\begin{pmatrix}
    b_{IT} \\
    b_{ES}
\end{pmatrix} \sim \text{MVN}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, D\right].
\]

(11)

where

\[
D = \begin{pmatrix}
    \sigma_b^2 & 0 \\
    0 & \sigma_b^2
\end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix}
    \sigma_{b,IT}^2 & 0 \\
    0 & \sigma_{b,ES}^2
\end{pmatrix},
\]

under $H_0$ and $H_1$, respectively.

The joint shared Gamma frailty model is formulated in the following way. Under the null hypothesis, the heterogeneity parameter $\theta$ is shared by the countries, while, under the alternative, the parameter is country-specific. Note that, under the two alternatives, the cumulative FOI remains country/disease-specific as before. Hence the probability of being susceptible for both infections under the alternative can be rewritten as
\[ \pi_{00}(d_i) = \begin{cases} \left[ \exp\left( \frac{\Lambda_{1,IT}(d_i)}{\theta_{IT}} \right) + \exp\left( \frac{\Lambda_{2,IT}(d_i)}{\theta_{IT}} \right) - 1 \right]^{-\theta_{IT}} & I_i = 1, \\ \left[ \exp\left( \frac{\Lambda_{1,ES}(d_i)}{\theta_{ES}} \right) + \exp\left( \frac{\Lambda_{2,ES}(d_i)}{\theta_{ES}} \right) - 1 \right]^{-\theta_{ES}} & I_i = 0. \end{cases} \]  

(12)

Note that under the null hypothesis \( \theta_{IT} = \theta_{ES} = \theta \). The remaining three probabilities are defined in a similar way.

Table 3 and 4 present the parameter estimates for the variance of random effects, goodness-of-fit measures (AIC and BIC), and the result of LRT for the three link functions, for GLMM and shared Gamma frailty model, respectively. For both random-effects models, regardless of the link function, the null hypothesis cannot be rejected, therefore we conclude that the level of individual heterogeneity in the countries is the same.

<table>
<thead>
<tr>
<th>Model</th>
<th>Country</th>
<th>( \sigma^2_b ) (95% C.I.)</th>
<th>LRT c.v.</th>
<th>AIC/BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weibull</td>
<td>Italy</td>
<td>0.78 (0.17, 1.93)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spain</td>
<td>0.60 (0.20, 1.24)</td>
<td>0.14</td>
<td>AIC=2490</td>
</tr>
<tr>
<td></td>
<td>IT &amp; ES</td>
<td>0.65 (0.30, 1.18)</td>
<td>((P = 0.71))</td>
<td>BIC=2527</td>
</tr>
<tr>
<td>Log-logistic</td>
<td>Italy</td>
<td>1.39 (0.27, 3.33)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spain</td>
<td>1.15 (0.31, 2.43)</td>
<td>0.07</td>
<td>AIC=2486</td>
</tr>
<tr>
<td></td>
<td>IT &amp; ES</td>
<td>1.24 (0.52, 2.24)</td>
<td>((P = 0.79))</td>
<td>BIC=2523</td>
</tr>
<tr>
<td>Log-normal</td>
<td>Italy</td>
<td>0.32 (0.03, 0.78)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spain</td>
<td>0.36 (0.10, 0.74)</td>
<td>0.03</td>
<td>AIC=2486</td>
</tr>
<tr>
<td></td>
<td>IT &amp; ES</td>
<td>0.34 (0.14, 0.62)</td>
<td>((P = 0.86))</td>
<td>BIC=2523</td>
</tr>
</tbody>
</table>

Table 3: Parameter estimates for the \( \sigma^2_b \) and 95% profile-likelihood CI for Italy, for Spain, and for the joint model; LRT for testing for common variance; information criteria for the joint model.

Figure 2 shows the estimated prevalences obtained from the best models for each family. Note that for the random-effects models the marginalized prevalence (discussed in Section 3.2) obtained from the models with common variance is presented. As expected, we notice that the proportions of IDUs infected with HCV and HIV increase according to the length of injecting career. In both countries, the prevalence of HCV infection increases very steeply until about 5 years of injecting drug use, while afterwards it levels out. In contrast, the prevalence of HIV infection presents a more linear profile over the years of injecting drug use, with different slopes for Italy and in Spain. Finally, we notice that the prevalence in Spain is generally higher than in Italy for both infections: averaging the prevalences obtained from the five models, we found that, after 10 years of injecting, in Italy the estimated HCV infection prevalence equals 77\% and the estimated HIV infection prevalence equals 6\%, whereas in Spain the respective values are 87\% and 36\%.
Table 4: Parameter estimates of the heterogeneity parameters $\theta$ and 95% profile-likelihood CI for the shared Gamma frailty models for Italy, for Spain, and for the joint model; LRT for testing for common variance; information criteria for the joint model.

<table>
<thead>
<tr>
<th>Model</th>
<th>Country</th>
<th>$\theta$ (95% C.I.)</th>
<th>$\sigma^2_\theta$ (95% C.I.)</th>
<th>LRT c.v.</th>
<th>AIC/BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weibull</td>
<td>Italy</td>
<td>2.59 (1.07, 176.43)</td>
<td>0.39 (0.01, 0.94)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spain</td>
<td>2.70 (1.38, 10.71)</td>
<td>0.37 (0.09, 0.73)</td>
<td>0.006</td>
<td>AIC= 2483</td>
</tr>
<tr>
<td></td>
<td>IT &amp; ES</td>
<td>2.68 (1.51, 7.03)</td>
<td>0.37 (0.14, 0.66)</td>
<td>(P = 0.94)</td>
<td>BIC= 2529</td>
</tr>
<tr>
<td>Log-logistic</td>
<td>Italy</td>
<td>2.39 (1.01, 52.11)</td>
<td>0.42 (0.02, 0.99)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spain</td>
<td>2.52 (1.31, 8.93)</td>
<td>0.40 (0.11, 0.76)</td>
<td>0.006</td>
<td>AIC=2490</td>
</tr>
<tr>
<td></td>
<td>IT &amp; ES</td>
<td>2.48 (1.43, 6.15)</td>
<td>0.40 (0.16, 0.70)</td>
<td>(P = 0.94)</td>
<td>BIC=2538</td>
</tr>
<tr>
<td>Log-normal</td>
<td>Italy</td>
<td>2.76 (1.15, 102.81)</td>
<td>0.36 (0.01, 0.87)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spain</td>
<td>2.99 (1.57, 10.12)</td>
<td>0.33 (0.1, 0.64)</td>
<td>0.013</td>
<td>AIC=2491</td>
</tr>
<tr>
<td></td>
<td>IT &amp; ES</td>
<td>2.92 (1.69, 7.18)</td>
<td>0.34 (0.14, 0.59)</td>
<td>(P = 0.91)</td>
<td>BIC=2539</td>
</tr>
</tbody>
</table>

4.3 Modeling the Association between HCV and HIV Infections as Function of Behavioral Risk Factors Using Marginal Models

We discuss now the results of the marginal models, when we let the association depend on the risk factors. The initial mean structure of the prevalence models include infection-specific intercepts and slopes for the logarithm of the length of injecting career, $\log(d_i)$, as well as infection-specific coefficients for the risk factors used in turn for the association model. Then, the mean structure is simplified based on the Wald’s test for coefficients.

We first examine the results for the continuous behavioral risk factors, namely, the length of injecting career and the age at first injection. Note that this information is available for both Italy and Spain. Figure 3 shows the odds ratios $\psi$ from ALR and BDM, while Figure 4 shows the correlation coefficients $\rho$ obtained from BPM. The 95% pointwise CI for the odds ratios from ALR are asymptotic, while the intervals for the odds ratios from BDM and for the correlations from BPM are 95% studentized bootstrap CI (Davison and Hinkley, 1997), based on $B = 999$ bootstrap replicates.
Figure 2: Observed (points) and estimated (lines) prevalence for HCV infection (upper curve) and for HIV infection (lower curve) for Italy and Spain. The fitted models are the ALR (black solid lines), the BDM (red dashed lines), the BPM (green dotted lines), the shared Gamma frailty model with common variance (blue dash-and-dot lines), and the GLMM with common variance (purple dashed lines). The area of the points is proportional to the sample size per length of the injecting career group.
Table 5: Parameter estimates for the association measures (OR for ALR and BDM, ρ for BPM) across the levels of the behavioral risk factors. For ALR we report 95% asymptotic CI, whereas for BDM and BPM we report 95% bootstrap studentized CI. The results presented in the table have been obtained from a complete case analysis.

Figures 3 and 4 (panels a and b) show the estimated functions for the odds ratio and the correlation for the length of injecting career. The final mean structure of the prevalence contains infection-specific intercepts and slopes for the length of injecting career. The two types of association generally rise with the length of injecting career in both countries. Based on the estimates of the association measures with their 95% CI, when injecting less than two years, odds ratios are lower than 1 and the estimated correlation is negative, implying that it is difficult to be already co-infected by HCV and HIV in the first years of injecting. However, after already two years of injecting, the estimated correlation becomes positive and the odds ratio larger than one. Hence, with increasing years of injecting, being infected with only one virus becomes less common and individuals tend to be increasingly infected with both viruses or not to be infected at all.

The results obtained for the age at first injection are shown in panel c and d of Figures 3 and 4. The final mean structure of the prevalence models includes the infection-specific intercepts and slopes, but not the risk factor, because it is not significant: indeed, the P-values for this risk factor from the ALR are 0.13 and 0.38 in Italy and Spain, respectively. As regards the association measures, the results differ between the two countries. In Spain both association measures increase with IDU’s age at first injection and the 95% pointwise CI never include the value of independence (1 for the odds ratio and 0 for the correlation); instead, in Italy we found that the odds ratio computed with ALR and the correlation are not significant, while the odds ratio computed with BDM is significant.

We turn now to examine the results for the categorical behavioral risk factors, i.e., sharing syringes on a ever/never basis and frequency of current injecting, information only available for Spain. The parameter estimates for the association measures are shown in Table 5. Note that the results presented in this table have been obtained from a complete case analysis.
For sharing of syringes, the final mean structure for the prevalences includes infection-specific intercepts and slopes for $\log(d_i)$ and sharing syringes as additional covariate. As expected from the literature (Mathéi et al., 2006), sharing syringes is a significant behavioral risk factor for both infections. This indicates that, adjusting for the length of the injecting career, the prevalence of HCV and HIV infections among IDUs never sharing syringes is significantly lower than among IDUs ever sharing syringes. However, the odds ratios are significantly greater than 1 and the correlation coefficients are significantly positive only for IDUs who have never shared syringes (see Table 5).

Finally, the effect of frequency of current injecting is not significant in the prevalence models; in the association model, the association parameters are significant only for current non-IDUs (persons who have not injected in the last month), as we can see from BDM and BPM, but not from ALR.

5 Conclusions

In this paper, we presented a general statistical framework to study the association between multiple infections, and we applied it to the case of co-infection with HCV and HIV in IDUs. The marginal models (ALR, BDM, and BPM) allowed us to estimate association measures between HCV and HIV infection, i.e., odds ratios and correlation coefficients. We have shown that the length of injecting career, the age at first injection, ever sharing of syringes, and, at a lesser extent, the frequency of current injecting are behavioral risk factors which may influence either the prevalence of HCV and HIV infection, their association or both.

The length of injecting career is known to be a main determinant of prevalence: indeed, we found that the prevalence of both infections increases with exposure time. In addition, we have shown that this risk factor affects the association between both infections as well. In particular, the odds ratio and the correlation slightly increase with the length of injecting use, indicating that people with a longer history of injecting have higher odds of experiencing both infections (or none), as opposed to having just one infection, or, in other words, the two infections become more correlated at the individual level.

The age at first injection is not a significant behavioral risk factor for the prevalence of the infections, but has a significant positive effect on the association measures, at least in Spain: older beginners have lower odds of experiencing just one infection and show a stronger correlation between both infections than younger beginners.
Figure 3: Odds ratio $\psi$ between HCV and HIV infection depending on the length of the injecting career (upper part) and on the age at first injection (lower part), obtained from BDM Weibull with 95% pointwise bootstrap studentized CI ($B = 999$ replicates, black dashed lines), and from ALR log-logistic with 95% pointwise asymptotic CI (red dotted lines). The solid horizontal line is equal to $\psi = 1$, where the association is not statistically significant.
Figure 4: Correlation $\rho$ between HCV and HIV infection depending on the length of injecting career (upper part) and on the age at first injection (lower part), obtained from BPM with 95% pointwise bootstrap studentized CI ($B = 999$ replicates, dashed lines). The solid horizontal line is equal to $\rho = 0$, where the association is not statistically significant.
We have shown that ever sharing syringes is a significant behavioral risk factor for the prevalence of both infections; in particular, IDUs who admit ever sharing syringes are characterized by higher prevalence for HCV and HIV infections. However, the association between these infections (as measured by the odds ratio and the correlation coefficient) is significant only in the group of IDUs who report having never shared syringes. One possible explanation of this finding can be found in the concept of individual heterogeneity (Coutinho et al., 1999; Farrington et al., 2001). Sharing syringes implies more homogeneous mixing of IDUs and thus more homogeneous transmission of viruses among individuals. The lack of association in those who have ever shared syringes is due to a balance between concordant statuses (mostly co-infection) and discordant ones (one infection, mostly with HCV). For this group of IDUs, the joint distribution of the infection statuses of HCV and HIV is not statistically different from the joint distribution of the infection statuses under independence. On the contrary, in the group of IDUs who report having never shared syringes, there is a preponderance of subjects with concordant statuses (mostly neither of the infections) over those with discordant ones (one infection, again mostly with HCV). A possible reason for the observed individual heterogeneity in the association of HIV and HCV infections in those IDUs who report having never shared syringes may be that this group consists of subgroups differing strongly in their actual behavioral risk, one of them being the group who truthfully reports having never shared, while the other one is made of those IDUs who do not admit their sharing behavior. In addition, these differences could be due to some unobserved effects, such as paraphernalia sharing and unhygienic injecting in general, which have been seen as possible transmission routes for HCV infection, but not for HIV infection (Lucidarme et al., 2004; Mathëi et al., 2006).

Finally, we found the effect of the frequency of current injecting on the association to be not very strong, except for the group of non-current injectors, who show higher odds ratios and correlations between HCV and HIV infections. As well for sharing syringes, this effect may be explained in terms of individual heterogeneity. A higher frequency of injecting, possibly related to the sharing of syringes, may imply more homogeneous mixing of IDUs and thus more homogeneous transmission of infections among IDUs, meaning that the probabilities of concordant and discordant statuses are similar. In contrast, the group of non-current injectors shows more heterogeneous mixing. This could be due to the merging of IDUs who gave up injecting with different timing, thus the probability of concordant statuses is higher than the probability of discordant ones.

The second group of models discussed in the paper deals more directly with the issue of individual heterogeneity by including subject-specific random effects in the model. For both GLMM and shared Gamma frailty model, a test for the variance of random effects is used to assess the degree of heterogeneity in the way
IDUs acquire the infections. We have shown that, for both models, subject-specific random effects are needed, indicating that, in Italy and in Spain, HCV and HIV infection statuses within IDUs are correlated. Furthermore, we have shown that the null hypothesis of common variance in Italy and in Spain cannot be rejected: this means that the same level of individual heterogeneity can be assumed for both countries.

The study presents some limitations related to the data. It may be that, to a certain extent, national differences are confounded by differences in study methods. First, we consider that in Italy the sampled subjects were IDUs who self-selected for a treatment in a DTC, while in Spain the sampled subjects were street-recruited IDUs. Second, in Spain only drug users younger than 30 years were included, while in Italy there was no age limit. Notwithstanding these differences, the authors of both surveys recognized that the prevalence results of the surveys were in line with other national prevalence results and that similar prevalence rates had been reported in Spain, Italy, Portugal and France (Camoni et al., 2010; Barrio et al., 2007).

In this paper we have used marginal models and random-effects models to estimate both prevalence and association measures. The marginal models focus on population-averaged quantities, such as the prevalence in the population or the association measures (odds ratios, correlation coefficients). When subject-specific questions are of interest, such as the individual heterogeneity in the population, GLMMs and shared Gamma frailty models ought to be considered, because of their use of individual-specific random effects. As shown in Section 3.2, since such conditional models allow derivation of population-averaged prevalence, these models can also be used for the same purpose of the marginal models. However, an advantage of the marginal models in such a setting is that the effects of direct interest (population prevalence and association measures) are captured directly by model parameters or simple functions thereof, unlike GLMMs or shared Gamma frailty models.

References


DOI: 10.2202/1948-4690.1010


