University of Tartu

Faculty of Science and Technology

Institute of Mathematics and Statistics

Jürgen Rannap

Causal effect of incarceration history on HIV susceptibility among people who inject drugs: an individual-level meta-analysis

Mathematics and Statistics

Mathematical Statistics

Master's Thesis (30 ECTS)

Supervisors: Prof Krista Fischer
Prof Anneli Uusküla
MSc Sven E. Ojavee

Tartu 2022

CAUSAL EFFECT OF INCARCERATION HISTORY ON HIV SUSCEPTIBILITY AMONG PEOPLE WHO INJECT DRUGS: AN INDIVIDUAL-LEVEL META-ANALYSIS

Master thesis

Jürgen Rannap

Abstract

The objective of this master's thesis is to describe the causal effect of ever being imprisoned on the probability of having human immunodeficiency virus among people who inject drugs. Here, we meta-analyse individual-level patient data that has been collected throughout Europe. Firstly, we provide an overview of the individual patient data and the process of merging the data sets into a complete analysis data set. Subsequently, we provide an overview of the bias reduction and causal inference and modelling methods such as propensity score matching and generalised linear mixed model. Finally, the thesis focuses on applying these practices on the merged data set. As a result, the thesis confirms the existence of a risk-increasing effect of ever imprisonment to the probability of having human immunodeficiency virus.

CERCS research specialisation: P160 Statistics, operations research, programming, financial and actuarial mathematics.

Key Words: Statistical inference, causality, meta-analysis, imprisonment, HIV.

VANGISTUSAJALOO PÕHJUSLIK MÕJU HIV TÕENÄOSUSELE SÜSTIVATE UIMASTIKASUTAJATE SEAS: INDIVIIDI-TASANDI METAANALÜÜS

Magistritöö

Jürgen Rannap

Lühikokkuvõte

Käesoleva magistritöö eesmärk on kirjeldada vangistusajaloo ja inimese immuunpuudulikkuse viirusesse sattumise põhjuslikku seost süstivate uimastikasutajate seas kasutades selleks üleeuroopaliselt kogutud individuaalse patsiendi andmete metaanalüüsi. Esmalt tutvustame antud vaatlusuuringu jaoks kogutud andmeid ja andmete terviklikuks analüüsiandmestikuks koondamise protsessi. Järgnevalt anname ülevaate analüüsiks kasutatavatest nihke vähendamise ja põhjusliku seose hindamise tehnikatest ja modelleerimismeetoditest nagu kalduvusskoori sobitamine ning üldistatud lineaarsed segamudelid. Lõputöö viimane osa keskendub tutvustatud meetodite rakendamisele analüüsiandmestiku peal. Töö tulemus kinnitab vangistusajaloo riski suurendava mõju olemasolu inimese immuunpuudulikkuse viirusesse sattumise tõenäosusele.

CERCS teaduseriala: P160 Statistika, operatsioonianalüüs, programmeerimine, finants- ja kindlustusmatemaatika.

Märksõnad: Statistiline järeldamine, põhjuslikkus, metaanalüüs, vangistus, HIV.

Contents

Introduction

Observational trials, focusing on causal estimations, suffer greatly from the fundamental problem of causal inference, which states that only one potential outcome under an exposure is ever measured. Randomised controlled trials overcome this by randomly assigning the exposure in trial subjects. In observational studies, however, the challenge arises from the outset of the trial and is fundamentally embedded in the design, where exposure groups are generally not randomly assigned. This introduces several obstacles in estimating accurate causal relationships, including confounding bias along with selection bias. To overcome these obstacles, several bias reduction techniques have been developed including methods utilising the propensity score, formulated by P. R. Rosenbaum and D. B. Rubin (Rosenbaum and Rubin, 1983).

Here, we investigate the causal effect of ever imprisonment on Human Immunodeficiency Virus (HIV) among people who inject drugs (PWID) in Europe according to individual patient data collected in the project *"European Study Group for Mathematical Modelling and Epidemiological Analysis of Drug Related Infectious Diseases"*, coordinated by the *European Monitoring Centre for Drugs and Drug Addiction*. We apply propensity score techniques for bias reduction in this observational study of individual patient data meta-analysis. The thesis is a comprehensive overview of the article (Wiessing, Uusküla, and Rannap et al., 2021) submitted to *The Addiction*.

Injection drug use occurs in most countries, and infection with HIV and hepatitis C virus (HCV) is prevalent in many populations of PWID. PWID are at high risk for contracting HIV in relation to sharing needles, syringes, or other drug injection equipment (e.g., cookers) with HIV infected peers. Globally, about 20% of PWID are living with HIV infection (ranging from 1% in Australasia to 36% in Latin America) (Degenhardt et al., 2017). Also, PWID experience a high prevalence of incarceration as a result of the drug-related crimes. In addition, there is a substantial overlap of PWID and people who have been imprisoned in Europe, stretching from 5% in France to 50%

in Estonia (Kivimets et al., 2018). The behaviour increasing HIV risk among PWID does not only occur inside prisons. A systematic review concerning recently released people in the population of PWID suggest that those people have substantially higher risk of HIV infection (Stone et al., 2018).

Alternatively, imprisonment could be an opportunity for diagnosing HIV and providing relevant treatment and intervention including opioid substitution treatment, needle and syringe programmes and anti-retroviral therapy for PWID. To our knowledge, this study is the largest individual patient data meta-analysis conducted as yet, assessing the causal relationship between imprisonment and HIV among community-recruited PWID.

The thesis is divided into three sections. In the first section, we give an overview of the data collected and the unification of the data into the complete HIV data set. The second section introduces the theoretical framework later used in the analysis. The final section focuses on the application of the methods on the HIV data set to estimate the causal effect of HIV on imprisonment among PWID.

1 Data Description

The goal of this chapter is to describe how to transform the data to conduct individual-level meta-analysis. The data was collected from people who inject drugs (PWID) currently from 82 cross-sectional studies carried out across 13 countries in Europe from 2001 to 2016. These 82 distinct studies were further grouped into 22 sites according to the country, city and year of the study. The sites varied from the data collected, the sampling method used and the final received sample size.

Compared to the meta-analysis of aggregated summary data, individual-level meta-analysis uses combined subject level data of multiple studies to assess a certain research question. This allows to better control the exact model specification and handle missing data. Therefore, bias reduction and a better control over confounding can be achieved (Stewart and Tierney, 2002). In order to carry out the individual-level meta-analysis across all the sites, the data were unified as the definitions of variables differed across the sites.

1.1 Unification of data across sites

The unified variable definitions were created such that the maximum amount of information could be used, while retaining the integrity of the collected data. The data unification procedure across the 22 sites (Table 1.2) included the definitions for each of the variables of interest. As the data varied across the sites in the variable definitions or specific recall periods used, the definitions of these unified variables had to be delicately constructed.

Raw variables defined either as numeric or categorical depending on the site were mapped to either categorical or numeric types. If the unified variable definition was more accurate than the definition in the site, it was decided to logically replace the values not defined as accurately as proposed in our unified variable definitions. As a result, to transform categorised numeric values into numeric, the middle points of the categories were used as the corresponding numeric values. When transforming numeric values into categories,

naturally the corresponding categories to which the numeric value belongs to was chosen.

When defining the classes for categorical variables, new categories were derived by unifying intersecting categories. This resulted in the categories being defined by the largest boundaries possible to include the smaller categories within. Although this technique introduces some loss of information, it enables combining variables with different levels of precision.

In Table 1.1, the definitions of the unified variables used in the analysis are shown with their corresponding priority, variable type and categories. The variables Country, City and Year uniquely identify a site.

The main variables of interest were binary, describing whether the individual has been infected with HIV and whether the person has ever been imprisoned. The primary goal of the analysis was to describe the causal relationship between those two primary variables. Secondary variables were included to adjust the imprisonment-HIV analysis for the potential confounding effects and to understand the effects of the secondary variables on the probability of having HIV. The aggregate level variable measuring wealth inequality, defined on the site level, was included to further control for potential confounding. The primary variables were available with the same definitions in most of the sites with the exception of the Czech Republic (CZ) and Portugal (PT-P) data for HIV, where only the self-reported HIV status was available, whereas in other sites various test to detect HIV were performed. A logical imputation step of the primary variable HIV was performed using the information of subjects receiving antiretroviral therapy (ART) in studies, where this information was collected. That allowed missing HIV statuses to be replaced with having HIV in the cases where the subject had received ART.

The unification procedure also required defining joint recall periods. The recall periods were unified by taking firstly into account the information that was of interest to us in the variable definition, either showing the subject's habits of certain risk measures across the entire lifespan or of a shorter period, and secondly the availability of such information on the site level. For Impris-

Table 1.1: Variable definitions of the unified data

Variable name	Priority	Variable type	Definition	Categories
Country	ID variable	Categorical	The country of the site.	
City	ID variable	Categorical	The city of the site.	
Year	ID variable	Categorical	The year of the data collection on site	
HIV	Primary	Categorical	Is the subject tested HIV positive in the current testing?	Y, N
Ever in prison	Primary	Categorical	Has the subject ever been imprisoned?	Y, N
Sex	Secondary	Categorical	The sex of the subject.	M, F
Age	Secondary	Numeric	The age of the subject in years.	
Duration of injecting	Secondary	Numeric	Duration of injection in years.	
Frequency of injecting	Secondary	Categorical	Frequency of injection, the most recent data.	Daily or more, Less than daily
Recently sharing syringes	Secondary	Categorical	Has the subject shared needle or syringe recently?	Y, N
Ever sharing syringes	Secondary	Categorical	Has the subject ever shared needle or syringe?	Y, N
Categorical number of sex partners	Secondary	Categorical	The number of sex partners in the last year.	0, 1, 2-9, $\geq$10
Number of sex partners	Secondary	Numeric	The number of sex partners in the last year.	
Main source of clean syringes	Secondary	Categorical	Primary source of clean syringes.	Pharmacy, NSP & outreach, Other
Opioid substitution therapy	Secondary	Categorical	Has the subject ever received opioid substitution treatment?	Y, N
Main drug injected	Secondary	Categorical	Primary substance(s) injected most recently.	Opioid, Cocaine, Opioid & Cocaine, Amphetamine, Other
Overdose	Secondary	Categorical	Has the subject ever experienced overdose?	Y, N
Gini index	Aggregated	Numeric	Wealth inequality measure, higher value indicates greater inequality	

onment (Ever in prison), needle or syringe sharing (Ever sharing syringes), Opioid substitution treatment and Overdose we used the entire lifespan of the subject as the recall period. The recall period of an entire lifespan was used as such variable definition is easily recollected by a subject and it shows the subject's habits across their entire lifespan.

The sharing of needles or syringes was categorised into two variables Recently sharing syringes and Ever sharing syringes. The first with an unfixed recall

time and the other with the recall time of a subject's lifespan. The recall period was chosen not to exceed one month as that was assumed to express the subject's recent behaviours regarding syringe or needle sharing. The number of sex partners was counted within a period of one year and for the respective categorical variable the numeric value was mapped into categories with the same recall period of one year.

The recall periods for Main drug injected and Main source of clean syringes varied the most across different sites. For these variables, it was decided to use the most recent recall period available from each site. Although this still left some discrepancies between the recall periods between sites, the majority of the recall periods are one month except for 6 and 12 month recall periods.

Aggregate-level measures that could have an impact on the outcome were also assessed. The aggregate-level variables included: Opioid substitution therapy coverage of opioid users in community, Needle and Syringe Programme (NSP) coverage of PWID in community, HIV prevalence among PWID in community, PWID prevalence in community, Antiretroviral Therapy (ART) coverage among PWID in community, incarceration rate in general population, incarceration rate in PWID, Gini index, unemployment in general population, Opioid substitution therapy coverage in prisons, NSP in prisons, HIV prevalence in PWID in prisons, PWID prevalence in prisons, and condom coverage in prisons. However, Gini index was the only aggregate-level variable kept in the final analysis as other aggregate-level variables did not improve the model fit.

The unification procedure also left out a large number of variables collected within sites. These variables were not included into the final data set being either not beneficial in describing the causal relationship in question, having too many sites with the information not being collected or having the data be considerably of poor quality. Examples of this included the type of first drug ever used, prostitution (having sex for money or drugs), alcohol use habits or disorder, smoking habits and nicotine use, etc.

Using the unified variable definitions, the site-specific data was merged. Concatenating the data from different sites left us with a total of $n = 43,807$

subjects in the final data set with sample sizes ranging from $n = 95$ in site PL-W to $n = 27,823$ in site UK-WEnI. Table 1.2 gives an overview of the number of subjects, the distribution of primary variables of interest and missing data information on these variables by each site.

1.2 Missing data

All of the data that was not available from the sites was initially coded as "not available" (NA). The missingness in the data set was bifold, containing missing data on the subject level and complete missingness of a variable on the site level, the latter of which occurring when the information was not collected from the sites. The amount of missing data was substantial, with the number of observations with no missing values being only 418 (0.95%). Figure 1.1 gives an overview of the missingness across the entire data set, showing the scope of missing data points by variable. The primary variables of interest HIV and Ever in prison have missingness of 1.9% and 3.3% respectively while variables including Recently sharing syringes, Main source of clean syringes and Overdose all have missingness percentages of over 50%. Furthermore, as part of missingness came from the sites not collecting the information, the amount of missing data across sites and therefore countries varied greatly (see Appendix 1).

The large proportion of missing observations in the data implied the necessity to take missingness into account and avoid potentially biased inference. In the analyses, we assumed that the missing data were missing at random as most of the missingness occurred because the information was not collected by the sites.

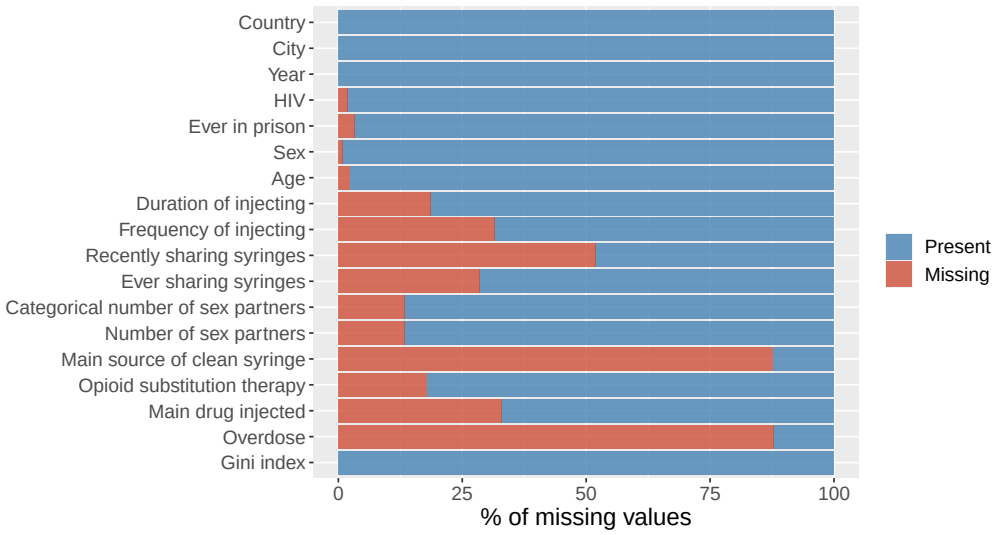


Figure 1.1: Proportion of missing (red) and observed (blue) data points in the combined data set.

Table 1.2: Sample size and distribution of primary variables of interest by site. Number of missing data shown in parentheses

Site	**Sample size**	**Sex**$_{\text{male}}$	**HIV**$_{+}$	**Imprisoned**
CZ	760	65.1%	0.23% *(332)*	39.8% *(2)*
EE-T	1031	79.1%	52.3%	62.4%
FI-7	600	68.4% *(14)*	1.19% *(12)*	34.9% *(19)*
GR-A	3320	84.5% *(1)*	15.2%	48.8% *(13)*
HU	1054	74.4%	0.19%	49.2% *(7)*
LV-5s	666	66.9% *(14)*	11.0% *(319)*	45.7% *(12)*
LV-R	290	74.5%	31.0%	50.3%
LU	420	75.2% *(5)*	17.5% *(112)*	40% *(130)*
NL-A	262	67.6%	11.5% *(18)*	89.1% *(14)*
PL-G	200	77.0%	25.3% *(2)*	33.7% *(1)*
PL-GK	193	75.1%	18.9% *(8)*	50.8% *(4)*
PL-Ms	776	71.4% *(10)*	18.0% *(10)*	45.7% *(15)*
PL-W	95	74.5% *(1)*	14.7%	30.9% *(1)*
PT-P	253	86.6%	42.3%	40.9% *(1)*
RU-5s	520	68.5%	37.5%	51.2% *(4)*
RU-IN	593	71.8%	19.1%	17.5% *(5)*
RU-StP	811	77.8%	55.7%	33.8%
RU-V	310	81.9%	5.48%	15.2% *(1)*
SP-C	730	82.4% *(1)*	27.3% *(16)*	70.1%
SP-MBS	637	74.3%	25.3%	48.7%
UK-EWnI	27,823	73.6% *(331)*	1.21%	64.0% *(1208)*
UK-S	2463	69.7% *(14)*	0.77%	65.7% *(22)*
Total	43,807	74.3% *(391)*	7.21% *(829)*	58.7% *(1459)*

Note: CZ = Czech Republic; EE-T = Tallinn, Kohtla-Järve, Estonia; FI-7 = seven cities (Helsinki, Vantaa, Espoo, Tampere, Turku, Lahti, Hämeenlinna), Finland; GR-A = Athens, Greece; HU = Hungary; LV-5s = five geographical areas (Riga, Jurmala, Ogre, Liepaja, Bauska), Latvia; LV-R = Riga and surrounding areas, Latvia; LU = Luxembourg; NL-A = Amsterdam, Netherlands; PL-G = Gdańsk, Poland; PL-GK = Gdańsk, Kraków, Poland; PL-Ms = six regions: Mazowieckie (Warszawa), Lubuskie (Zielona Góra, Gorzów Wlkp., Cibórz, Nowy Dworek), Śląskie (Katowice, Chorzów, Sosnowiec), Dolnośląskie (Wrocław – 2 locations), Lubelskie (Lublin, Puławy), Warmińsko-Mazurskie (Olsztyn, Elbląg, Barczewo), Poland; PL-W = Warszawa, Poland; PT-P = Porto, Portugal; RU-5s = five cities (Barnaul, Volgograd, Naberezhnye, Chelny, Perm, Abakan), Russia; RU-IN = Ivanovo, Novosibirsk, Russia; RU-StP = Saint Petersburg, Russia; RU-V = Voronezh, Russia; SP-C = Catalonia, Spain; SP-MBS = Madrid, Barcelone, Seville, Spain; UK-EWnI = England, Wales & Northern Ireland, United Kingdom; UK-S = Scotland, United Kingdom

2 Methodology

In this chapter, we give an overview of the theoretical framework later used in the analysis. We describe the techniques of propensity score matching, widely used for the reduction of bias in the case of observational study data; multiple imputation, a flexible tool for dealing with missing data problems; and generalized linear mixed models, an extension of linear mixed models used to model correlated non-Gaussian responses.

2.1 Propensity score matching

Propensity score techniques are widely used in observational studies to reduce the effects of confounding and selection bias. Confounding arises from covariates affecting both exposure and the outcome when not adequately controlled, whilst selection bias is the result of a non-random selection process of the study sample, with subjects having disparate probabilities of being selected (Haneuse, 2016). Propensity score methods adjust for these discrepancies with the balancing of baseline covariate distributions in the exposed and unexposed treatment groups using a balancing score - propensity score (Rosenbaum and Rubin, 1983). In the case of randomised controlled trials, the balancing of baseline covariates is achieved with the study design (exposure randomisation) and a direct comparison of exposure groups is feasible for estimating the treatment effect. In observational trials, however, where generally such balance is not present, a crude comparison of exposure groups might not be viable (see Fundamental Problem of Causal Inference (Holland, 1986)) and therefore can be subject to considerable bias.

Propensity score matching, a propensity score based bias reduction algorithm for observational studies, tries to imitate a randomised sample by matching exposed and unexposed subjects on the basis of the propensity score to estimate the effect of the exposure on an outcome. We use propensity score matching to create a sample in which imprisoned subjects are comparable to non-imprisoned subjects on all baseline covariates, mimicking a randomised

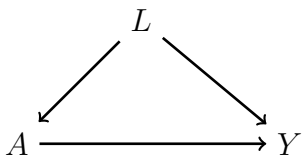


Figure 2.1: *Directed acyclic graph representing the causal relationships between a confounder set L, an exposure A and an outcome Y*

trial. We introduce the following definition for the propensity score, originally coined by Rosenbaum and Rubin.

Definition. (Rosenbaum and Rubin, 1983)
The propensity score $\pi(L)$ is the conditional probability that an individual is assigned to a particular exposure group given the set of observed covariates L.

Mathematically, given that L is a set of confounding covariates sufficient to adjust for confounding between a binary exposure A and an outcome Y, depicted as a causal diagram in Figure 2.1, the propensity score can be expressed as the following probability

$$\pi(L) = P(A = 1|L).$$

As the exposure variable A is binary, the propensity score can, for example, be modelled and computed by fitting a logistic regression model

$$\pi(L) = P(A = 1|L) = \frac{\exp(\alpha + \boldsymbol{\beta}'L)}{1 + \exp(\alpha + \boldsymbol{\beta}'L)},$$

where α is the model intercept, $\boldsymbol{\beta}$ is the parameter vector and L is the vector of confounding covariates.

To justify the reasoning behind the propensity score matching method, (Rosenbaum and Rubin, 1983) and (Hernán and Robins, 2020) have demonstrated two key properties of the propensity scores. Firstly, the equivalency of the propensity score to the balancing score which we later confirm, suggests that

grouping on similar propensity scores imitates a pseudorandomised trial given the set of observed confounders L. Secondly, that conditional exchangeability, given the confounders L, implies conditional exchangeability given the propensity scores $\pi(L)$. This is later used to show that instead of matching being conducted on the entire confounder set L, it is sufficient to conduct matching on the propensity score $\pi(L)$.

In the following section, we will consider A as a dichotomous exposure variable, Y as a dichotomous outcome variable and L as the set of confounding covariates sufficient to adjust for confounding between the binary exposure A and outcome Y. We start with the proof of the equivalency of propensity score to the balancing score.

Definition. (Rosenbaum and Rubin, 1983)
A balancing score $b(L)$ is a function of the observed covariates L such that the conditional distribution of L given $b(L)$ is the same for exposed $(A = 1)$ and unexposed $(A = 0)$. That is

$$L \perp\!\!\!\perp A|b(L).$$

We give the proof of the propensity score being a balancing score as a lemma below. For the proof, we use the law of total expectation.[1]

Lemma 1. *The propensity score $\pi(L)$ is a balancing score.*

Proof. The statement that the propensity score is a balancing score, by definition, means that given a propensity score $\pi(L)$

$$L \perp\!\!\!\perp A|\pi(L).$$

It is sufficient to show that

$$P(A = 1|\pi(L), L) = P(A = 1|\pi(L)),$$

[1]Let X and Y be random variables defined on the same probability space and let $E(X)$ be defined, then $E(X) = E(E(X|Y))$. Proof given in (Weiss, 2006).

which implies the conditional independence of L and A conditional on $\pi(L)$.

$$\begin{aligned} P(A=1|\pi(L),L) &= P(A=1|L) \\ &= \pi(L) \\ &= E(\pi(L)|\pi(L)) \\ &= E(E(A|L)|\pi(L)) \\ &= E(E(A|L,\pi(L))|\pi(L)) \\ &= E(A|\pi(L)) \\ &= P(A=1|\pi(L)). \end{aligned} \tag{1}$$

□

In the case of randomised trials, the study design prevents treatment assignment to be associated with any covariate. Lemma 1 ensures the same disassociation for observational trials, conditional on the propensity score $\pi(L)$. We can therefore conclude that given the set of observed confounders L, conditioning on the propensity score $\pi(L)$ imitates a pseudorandomised trial as the exposure and the set of confounder variables are independent conditional on the propensity score.

We proceed with the second key property of the propensity score, which indicates that conditional exchangeability transfers to the case of the propensity score. For showing this we start with the definition of counterfactuals.

Definition. (Hernán and Robins, 2020)
Let Y^a be the outcome variable that would have been observed under the treatment value $A = a$, $a \in \{0, 1\}$. The variable Y^a is referred to as a counterfactual outcome.

The counterfactual outcomes are widely used in the field of causal inference, as they allow for a simple formulation of the theory, being the main variable of interest. We continue with the definition of exchangeability using the counterfactual outcomes.

Definition. (Hernán and Robins, 2020)
If the counterfactual outcome Y^a and the treatment assignment A are independent for all $a \in A$, then we say that the exposed and unexposed subjects are exchangeable. That is

$$Y^a \perp\!\!\!\perp A, a \in A.$$

Conditional exchangeability, conditional on the covariate vector L, is now defined as

$$Y^a \perp\!\!\!\perp A|L, a \in A.$$

From the definition we can derive that under conditional exchangeability there are no unobserved differences in the outcomes between the exposure groups, had the two groups both received treatment within levels of the covariate L. The next lemma links the ideas of conditional exchangeability and propensity models, where exchangeability of the exposed and unexposed within levels of the covariates L indicates exchangeability of the exposed and unexposed within levels of the propensity score $\pi(L)$.

Lemma 2. *If the treatment assignment A is exchangeable within levels of the confounders L, then it is exchangeable within levels of the propensity score $\pi(L)$. That is for $a = 0, 1$*

$$Y^a \perp\!\!\!\perp A|L \Rightarrow Y^a \perp\!\!\!\perp A|\pi(L).$$

Proof. Let us assume treatment assignment A to be exchangeable within levels of the confounder set L, meaning

$$Y^a \perp\!\!\!\perp A|L, a \in A.$$

It now suffices to show that

$$P(A = 1|Y^a, \pi(L)) = P(A = 1|\pi(L)),$$

which implies the conditional independence of Y^a and A conditional on $\pi(L)$.

$$
\begin{aligned}
P(A = 1|Y^a, \pi(L)) &= E[A|Y^a, \pi(L)] \\
&= E[E(A|Y^a, L, \pi(L))|Y^a, \pi(L)] \\
&= E[E(A|Y^a, L)|Y^a, \pi(L)] \\
&= E(\pi(L)|Y^a, \pi(L)] \\
&= \pi(L) \\
&\overset{(1)}{=} P(A = 1|\pi(L)).
\end{aligned}
$$

□

Lemma 2 indicates that instead of matching being conducted on the entire confounder set L, it suffices to match on the propensity score $\pi(L)$. This follows straight from the lemma, as when conditional exchangeability holds, there are no unobserved differences in the outcomes between the exposure groups, had the two groups both received treatment given the propensity score $\pi(L)$. Furthermore, it can be shown that under conditional exchangeability the unbiased causal effect of an exposure on an outcome can be derived as a simple mean difference of the exposure groups within levels of the propensity score (Stuart, 2010). Therefore, the straightforward nature of the estimates justifies the implementation of propensity score matching in the case of observational study data, where the main assumption to be considered is the assumption of no unmeasured confounding, meaning that we account for all confounding in the estimation of the propensity scores.

2.1.1 Estimation of the propensity score

In randomised controlled trials, the true value of the propensity score is given by the design and is equal to a fixed probability, mostly 0.5, meaning the probability of an exposure conditional on the confounding variables is an even chance event. In observational studies, the true value cannot be obtained as the probability of exposure assignment is unknown. Therefore, an

estimate is needed of the true propensity scores (Austin, 2011). The propensity scores can be estimated using a logistic regression model with logit or probit link controlling for all the baseline covariates L (Austin, 2011; Hernán and Robins, 2020). Higher order terms and interactions between covariates can be introduced in the model to better the covariate balance between the exposure groups (Dehejia and Wahba, 1999). Propensity score estimation with the use of bagging and boosting, decision trees and random forests have also been investigated (Lee, Lessler, and Stuart, 2010).

In the case of variable selection to the propensity score model, it is crucial, to meet the assumption of conditional exchangeability, to include all variables known to be related to exposure assignment and outcome in the estimation of the propensity scores (Rubin and Thomas, 1996). When matching on the propensity scores, the inclusion of variables unrelated to the exposure assignment will have little influence on the propensity score model. On the other hand, the exclusion of a potentially important covariate can lead to a further increase in bias (Stuart, 2010). Therefore, regarding variable selection to the propensity score model, a liberal approach should be favoured to include variables with a potential association with the exposure assignment or the outcome.

2.1.2 Matching on the propensity score

Several propensity score techniques for bias reduction have been suggested, including covariate adjustment, inverse probability treatment weighting, subclassification and propensity score matching (Austin, 2011; Rosenbaum and Rubin, 1983). Here, we will focus on the method of matching on the propensity score. The matching algorithm requires the definition of a distance measure to assess the similarity between two different subjects - more precisely their propensity scores. The distance measure used later in the analysis is taken as the propensity score distance.

Definition. *Let π_i and π_j, $i \neq j$ be the propensity scores of subjects i and j respectively. The propensity score distance D_{ij} between the two scores is*

given by a scalar difference:

$$D_{ij} = |\pi_i - \pi_j|,$$

where $\pi_i = P(A_i = 1|L_i)$.

The matching procedure was implemented using a $k : 1$ nearest neighbour matching algorithm, as this is the most common design for matching on the propensity score (Zakrison, Austin, and McCredie, 2018). The nearest neighbour procedure matches each exposed subject to the nearest unexposed k subjects conditional on the propensity score distance measure. The choice of the matching ratio is fundamentally a bias-variance trade-off (Stuart, 2010), as for $k : 1$ matching, smaller values of k will lead to closer matches and therefore lesser bias. On the other hand, larger k values lead to increased sample size but also to less accurate matches. Usually, the choice of $k = 1$ is selected which on the one hand presents simplicity but also can lead to a large amount of discarded unexposed subjects. The variable ratio algorithm (Stuart, 2010) was implemented within the nearest neighbour matching which allowed k to vary within each match. This is beneficial in the case where we have large amounts of close matches, as controlling for the loss of close matches can aid in the goodness of the matching procedure in general. Matching with replacement can also be implemented, where a certain unexposed subject can be matched to several exposed subjects. This is beneficial in avoiding additional bias in the case where the number of unexposed subjects compared to exposed subjects is small (Stuart, 2010).

The goodness of the matching procedure can be assessed by the evaluation of covariate balance in the two matched groups, where covariate balance is measured as the similarity of the empirical distributions of all covariates (Stuart, 2010). The most common numerical balance measures are the standardised difference of means and the ratio of variances between the exposed and unexposed groups. The standardised mean difference (smd) is defined as the absolute difference of means of the i'th covariate in the exposed L_i^e and unexposed L_i^u groups divided by the pooled standard deviation using

the standard deviation in the exposed σ_i^e and the unexposed σ_i^u, calculated as

$$\text{smd}_i = \frac{|\bar{L}_i^e - \bar{L}_i^u|}{\sqrt{\frac{(\sigma_i^e)^2 + (\sigma_i^u)^2}{2}}}. \tag{2}$$

The variance ratio (vr) is defined as the ratio of variances in the exposed and the unexposed groups of the i'th covariate, calculated as

$$\text{vr}_i = \frac{(\sigma_i^e)^2}{(\sigma_i^u)^2}. \tag{3}$$

The post-matched standardised differences of means less than 0.25 and variance ratios bounded by 0.5 and 2 for a given covariate L_i suggest a moderately suitable covariate balance (Rubin, 2001).

2.1.3 Analysis on the matched data

The matching methods are in nature not a causal inference estimation technique and generally, a regression analysis is implemented to lower the residual covariate imbalance between the exposure groups (Stuart, 2010). After the matching algorithm links exposed and unexposed subjects, pooled matched data is compiled which includes all exposed and unexposed subjects and the baseline covariates to be included in the regression model. The pooled data set is assumed to have achieved sufficient balance of the baseline covariates and the data should have same properties as if it came from a randomised trial. For the variable ratio matching algorithm, the addition of weights to the matched subjects are introduced (Stuart, 2010). These weights are later directly implemented in the model fitting process through the use of weighted regression. The weights are calculated as a proportion of the number of unexposed matched to each exposed subject.

2.2 Multiple imputation

Multiple imputation is a conventional approach for dealing with the problem of missing data. The concept is an extension on single imputation, where missing values are replaced several times instead of only once. This fundamental difference of multiplicity signifies the advantage of multiple imputation over single imputation: single imputation procedures are not capable of accounting for the sampling variability in the imputations (Rubin, 1987). The concept of multiple imputation involves multiple replacements to the missing data to receive several complete data sets by random draws from the posterior predictive distribution given the observed data. These separate and complete data sets are then individually analysed and the results combined into a single estimate.

The two main multiple imputation methods are joint modelling and multiple imputation by chained equations (MICE). The joint modelling techniques all assume variables in the data set to follow a joint model e.g. multivariate normal or log-linear (Buuren and Groothuis-Oudshoorn, 2011). The determining of this joint model however can in itself pose great difficulty. A more flexible approach is the MICE procedure, where a series of regression models are run on a variable-by-variable basis, conditional on other variables in the data set (Azur et al., 2011). This allows variables with different distributions to be modelled based on a univariate distribution and leaves aside the need for an assumption of a joint higher degree model. The MICE procedure assumes that conditional on other variables in the imputation, the missing data is missing at random (MAR), which implies that the probability of a value being missing is independent of the unobserved variables, conditional on the observed variables (Azur et al., 2011). The assumption of MAR is relevant for the imputation procedure to produce unbiased estimates (Rubin, 1987).

According to (Azur et al., 2011), the MICE process can be summarised in the following steps:

1. A mean imputation is implemented for each missing value;
2. The imputations are set back to missing for one chosen variable;

3. The existing values of the chosen variable are regressed on all/some other variables in the data set;

4. The missing values of the variable regressed on are replaced with the imputed values, based on the predictions of the fitted model, using a chosen method (described later);

5. Steps 2-4 are repeated for each variable with missing values. The imputation of all variables having missing values is considered as one iteration;

6. Steps 2-4 are repeated for a given number of iterations and the imputed values updated at each iteration.

The imputation process is a recurrent one, with the entire process begin repeated when the assigned number of iterations has been completed. This repetition produces several complete imputed datasets that are later individually analysed. To assess the goodness of the imputation (Buuren and Groothuis-Oudshoorn, 2011) have suggested that the by-variable plotted densities of observed and imputed data can be compared.

For the imputation procedure, several imputation methods have been suggested (Buuren and Groothuis-Oudshoorn, 2011). In this thesis, we will specify the methods later used in the analysis process, which include predictive mean matching, logistic regression and multinomial logistic regression. The predictive mean matching algorithm, introduced by (Rubin and Schenker, 1986; Little, 1988), can be considered as a hot-deck imputation method for all cases of quantitative variables, where the missing values are replaced with a similar observed value. The predictive mean matching algorithm (Buuren, 2012) randomly selects the observed value as the imputation value from a pool of observed subjects, where the model-based predicted means of the variable in question are closest to the predicted mean of the missing observation. The logistic regression and multinomial logistic regression are used in order to model dichotomous and polytomous variables respectively and are explicitly described in (Greene, 2003).

2.2.1 Parameter estimation

The parameter estimates from the complete data sets are obtained with a separate analysis performed on each data set individually. The different parameter estimates are thereafter combined using Rubin's rule (Rubin, 1987).

The following expressions are taken from (Rubin, 1987). Given M complete data sets, the multiple imputation estimate $\hat{\theta}^*$ of a parameter θ is given by

$$\hat{\theta}^* = \frac{1}{M} \sum_{m=1}^{M} \hat{\theta}_m, \tag{4}$$

where $\hat{\theta}_m, m = 1, \ldots, M$ is the model parameter estimate of the m'th data set. The formula shows that the multiple imputation estimate $\hat{\theta}^*$ is obtained by an averaging measure of the different model parameters.

The total variance estimate V_{Total} of the multiple imputation estimate $\hat{\theta}^*$ is given by

$$V_{Total} = \underbrace{\frac{1}{M} \sum_{m=1}^{M} \hat{\sigma}_m^2}_{\substack{\text{between} \\ \text{imputation} \\ \text{variance}}} + \left(\frac{M+1}{M} \right) \underbrace{\frac{1}{M-1} \sum_{m=1}^{M} (\hat{\theta}_m - \hat{\theta}^*)^2}_{\text{within imputation variance}}, \tag{5}$$

where $\hat{\sigma}_m^2, m = 1, \ldots, M$ is the variance of the estimate $\hat{\theta}_m$. The total variance can be viewed as a combination of both between imputation and within imputation variances.

2.3 Generalised linear mixed model

Generalised linear mixed models (GLMM) are a tool of regression analysis used for modelling a structure of correlated non-Gaussian responses. GLMM can be viewed as a generalization of the linear mixed models, where the distribution of the response is assumed normal (Jiang, 2007). With GLMM we allow the response to be of various distributions, such as binary. GLMM

can also be viewed as an extension of the generalised linear models, where only uncorrelated responses can be included (Jiang, 2007). With GLMM we permit responses to have some correlation structure. The correlation structure enables modelling of data where similarities in response measures could, for example, stem from responses coming from the study site or multiple measurements of even the same subject.

Let the observations y_i, $i = 1, \ldots, n$ represent the univariate response variables for n subjects and $\boldsymbol{x}_i$, $\boldsymbol{z}_i$ the p- and q-dimensional vectors of explanatory variables associated with the p fixed and q random effects respectively. Further, suppose that $\boldsymbol{\beta}$ is a p-dimensional vector of fixed effects and $\boldsymbol{\alpha}$ is a q-dimensional vector of random effects where the responses y_i are conditionally independent, conditional on $\boldsymbol{\alpha}$, with means (Breslow and Clayton, 1993)

$$E(y_i|\boldsymbol{\alpha}) = \mu_i, \quad i = 1, \ldots, n. \tag{6}$$

As with the generalised linear models, the conditional mean μ_i in GLMM is also related to the linear predictor η_i via a link function $g(\cdot)$ with inverse $h = g^{-1}$ as

$$g(\mu_i) = \eta_i, \quad i = 1, \ldots, n, \tag{7}$$

where $\eta_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{z}_i'\boldsymbol{\alpha}$ (Breslow and Clayton, 1993). For a generic form of the model, let us denote

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1' \\ \boldsymbol{x}_2' \\ \vdots \\ \boldsymbol{x}_n' \end{bmatrix}, \quad \boldsymbol{Z} = \begin{bmatrix} \boldsymbol{z}_1' \\ \boldsymbol{z}_2' \\ \vdots \\ \boldsymbol{z}_n' \end{bmatrix},$$

then from (6) and (7) the conditional mean satisfies

$$E(\boldsymbol{y}|\boldsymbol{\alpha}) = h(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\alpha}).$$

We assume that $\boldsymbol{\alpha}$ is from a multivariate normal distribution with mean 0 and covariance $\boldsymbol{D}(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the component specifying the distribution of the random effects (Breslow and Clayton, 1993). The general form of GLMM for the response $\boldsymbol{y}$ can be therefore written as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\alpha} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ and $\boldsymbol{\alpha} \sim N(\mathbf{0}, \boldsymbol{D}(\boldsymbol{\theta}))$.

For the logistic mixed model, used as the main analysis model, we would assume the binary responses $y_1, \ldots, y_n$ to be conditionally independent Bernoulli distributed with $p_i = P(y_i = 1|\boldsymbol{\alpha})$. Drawing on the reasoning above we would have

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \boldsymbol{x}_i'\beta + \boldsymbol{z}_i'\alpha,$$

with the link function $g(\cdot) = \text{logit}(\cdot), i = 1, \ldots, n$.

2.3.1 Model parameter estimation

The GLMM model parameters can be estimated using the maximum likelihood method based on the full marginal likelihood function $L(\boldsymbol{\beta}, \boldsymbol{\theta})$ (Raudenbush, Yang, and Yosef, 2000). From the definition of the likelihood function we can derive the full marginal likelihood $L(\boldsymbol{\beta}, \boldsymbol{\theta})$ as an integral of the marginal density, marginalising out the random effect parameter $\boldsymbol{\alpha}$

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}|\boldsymbol{y}) = \int f_{\boldsymbol{y}|\boldsymbol{\alpha}}(\boldsymbol{y}|\boldsymbol{\alpha}, \boldsymbol{\beta}) f_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}|\boldsymbol{\theta}) d\boldsymbol{\alpha}, \tag{8}$$

where $f_{\boldsymbol{y}|\boldsymbol{\alpha}}(\boldsymbol{y}|\boldsymbol{\alpha}, \boldsymbol{\beta})$ is the probability density of the response vector $\boldsymbol{y}$ and $f_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}|\boldsymbol{\theta})$ is the probability distribution with a parameter vector $\boldsymbol{\theta}$ (Raudenbush, Yang, and Yosef, 2000).

The maximum likelihood estimation requires the maximisation of equation (8), where in general the integration can not be evaluated due to the high dimensionality of the distribution of random effects (Raudenbush, Yang, and Yosef, 2000). To overcome this, a range of approximation processes of the integral have been introduced. A well-known approximation process is the Laplace approximation, which yields asymptotically unbiased estimates of the model parameters (Jiang, 2007). The Laplace method takes advantage of the Taylor expansion and the fact that the conditional distribution of the response with the random effect follows an exponential distribution family (Raudenbush, Yang, and Yosef, 2000). This allows approximating the integral kernel $f_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}|\boldsymbol{\theta})$ by an expression proportional to the Gaussian probability density and as a consequence enables the approximate computing of the full marginal likelihood function.

2.4 Derivation of the confidence interval for relative risk using the Delta method

The Delta method (Agresti, 2013) states that if an arbitrary statistic T_n is normally distributed with a mean θ and a standard error of $\sigma/\sqrt{n}$, with the subscript n denoting the statistics T_n dependence on the sample size n, simultaneously written as a convergence in distribution, denoted by

$$\sqrt{n}(T_n - \theta) \xrightarrow{D} N(0, \sigma^2),$$

then for any function g, where $g'(\theta)$ exists and is not equal to 0 the following convergence holds

$$\sqrt{n}(g(T_n) - g(\theta)) \xrightarrow{D} N(0, \sigma^2[g'(\theta)]^2).$$

We can take advantage of the aforementioned method for finding the confidence interval for a relative risk (RR) from an odds ratio (OR) by noting that the distribution of log odds ratio, denoted here by L, is approximately normal (Agresti, 2013)

$$L \sim N(\log(OR), \sigma^2).$$

We propose for the function $g(\cdot)$ the following

$$g(L) := \frac{\exp(L)}{1 - p_0 + p_0 \exp(L)},$$

which from (Grant, 2014) we know is the conversion function from an odds ratio to a relative risk for $L = log(OR)$, where p_0 is the is the baseline risk. Now, for finding the confidence interval for the relative risk we need to calculate the derivative of $g(L)$ with respect to L.

$$\begin{aligned} g'(L) &= \frac{\exp(L)[1 - p_0 + p_0 \exp(L)] - p_0 \exp(L) \cdot \exp(L)}{[1 - p_0 + p_0 \exp(L)]^2} \\ &= \frac{\exp(L)(1 - p_0)}{[1 - p_0 + p_0 \exp(L)]^2}. \end{aligned}$$

Therefore, we know from the Delta method that

$$g(L) \sim N[g(\log(OR)), \sigma^2(g'(log(OR)))^2].$$

Now, as $g(L)$ is normally distributed and from $g(log(OR)) = RR$ we can obtain the 95% confidence interval for the population mean RR, which can be calculated as

$$\left(\frac{\exp(\hat{L})}{1 - p_0 + p_0 \exp(\hat{L})} \pm z_{0.975} \cdot \text{se}(\hat{L}) \frac{\exp(\hat{L})(1 - p_0)}{[1 - p_0 + p_0 \exp(\hat{L})]^2} \right),$$

where $\hat{L}$ is the estimated $\log(OR)$, $z_{0.975}$ the 0.975 quantile of the standard normal distribution and $\text{se}(\hat{L})$ the standard error of the estimated $\log(OR)$.

3 Statistical analysis of the HIV dataset

In this thesis, a propensity score matched sample is used to estimate the effect of imprisonment on the probability of having HIV among people who inject drugs. The individual-level data was collected in the form of survey results, received from several sites ($n = 22$) across 13 countries in Europe, meaning that the analysis was an individual patient-level meta-analysis. The effect of imprisonment on HIV was estimated using generalised linear mixed models with random effects of country and city-year of the respective survey. Due to a large number of missing values in the collected data, mainly at site level, an imputation process was implemented on the whole data set.

3.1 Imputation

The imputation procedure was implemented due to the amount of missing data in the combined data set being substantial (99.05% of subjects contain missing values). The large number of subjects containing missing values resulted in a complete case analysis to be considered inapt. Therefore, assuming that the data was missing at random (MAR), a more appropriate method of multiple imputation (MI) was chosen as it enables using the entire sample.

As the missingness occurred directly on the patient level (some values missing within each site) and also on the site or country level (entire variable missing from the site), the MI procedure was implemented in two stages. Firstly, the MI procedure was applied within each country and thereafter on the whole data set, as overlooking the clustering of countries and performing a single-level imputation would be subject to additional bias (Buuren, 2012). The within-country imputation was conducted to take into account the country-specific similarities of subjects, while the second stage of the imputation over the whole data set was implemented to compensate for the variables that were not measured within countries. Both imputation stages were carried out using five imputations ($m = 5$) with 30 iterations for each imputation step which resulted in 25 unique imputed data sets. The methods

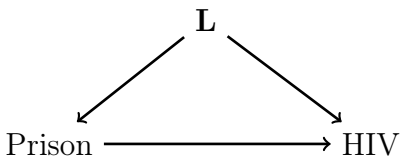


Figure 3.1: Directed acyclic graph representing the causal relationships between the set of confounders **L**, imprisonment and HIV.

used in all imputations were predictive mean matching for all quantitative variables, logistic regression method for binary variables, and multinomial logistic regression for imputing variables with more than two categories. The MI algorithm was implemented in R (version 3.6.1) using the package MICE (Buuren and Groothuis-Oudshoorn, 2011).

3.2 Propensity score matching

We used propensity score matching to construct a pseudo-randomised sample by matching imprisoned subjects to non-imprisoned subjects in order to estimate the effect of imprisonment on the probability of having HIV. For the propensity score matching algorithm, we followed the causal diagram given in Figure 3.1. In the diagram, we have the exposure of a patient having ever been imprisoned, assigned to the variable "Prison", the outcome of a patient having HIV, assigned to the variable "HIV" and the set of confounders **L**, which included all individual-level variables given in Table 1.1. To estimate the effect of imprisonment on being HIV positive, we tried to take the confounding between the confounder set **L** and imprisonment into account with the aid of propensity score matching, effectively erasing this confounding relationship. For the following sections, the city of the site and the year of data collection on site were combined into a joint variable "city-year".

3.2.1 Propensity score estimation

The propensity score, a probability of a PWID being ever imprisoned, can be estimated by fitting a regression model on imprisonment and is therefore highly dependent on the variable selection into the model. To acquire the model with the propensity score estimations closest to their true values and to best try to meet the assumption of conditional exchangeability, we have to control for all covariates which are associated with either imprisonment or having HIV.

The propensity scores were estimated using a logistic regression model with logit link, including all individual level variables given in Table 1.1 excluding HIV and with City and Year combined into city-year. To improve model fit and predictive power of the propensity score model, higher degrees and interactions of covariates were introduced. The final model is given in equation 9.

$$\ln\left[\frac{P\left("\text{Prison}" = 1\right)}{1 - P\left("\text{Prison}" = 1\right)}\right] = \beta_0 + \boldsymbol{\beta}\boldsymbol{L_1}.^{2} \tag{9}$$

The selection of the higher orders and interactions of covariates into the propensity score model was determined with different information criteria, including the Akaike and Bayesian information criteria. The distributions of the obtained propensity scores by the exposure imprisonment for all 25 data sets is given in Figure 3.2. The figure shows a relatively large intersection of the propensity scores by the imprisoned and non-imprisoned subjects. This overlap is a good indicator for the following matching procedure, where imprisoned subjects are matched to non-imprisoned subjects conditional on their propensity score.

A logistic mixed-effect model with logit link, where the variables city-year and Country were chosen as random effects, was also applied for the estima-

[2]Here β_0 is the intercept, $\boldsymbol{\beta}$ is the parameter vector and $\boldsymbol{L_1}$ the set of individual level confounders additionally including higher order terms of Age, Duration of injecting and Number of sex partners and the following interactions: city-year and (Duration of injecting, Age), Duration of injecting and (Sex, Age, Main drug injected) and Main drug injected and (Sex, Frequency of injecting).

tions of the propensity scores but later discarded due to lack of significant improvement and a major increase in computing time over the simpler logistic regression model.

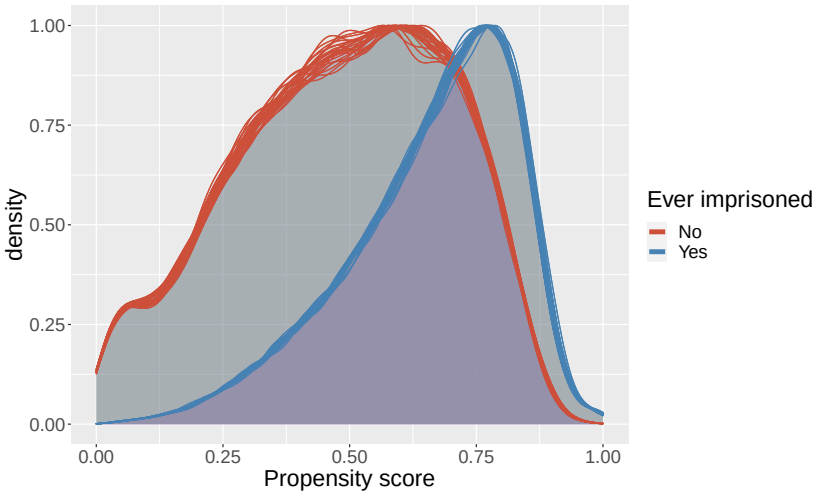


Figure 3.2: Distribution of propensity scores by ever imprisonment for all 25 complete data sets. Non-imprisoned subjects are represented with red, imprisoned subjects with blue curves.

3.2.2 Matching procedure

The aim of the matching procedure is to construct the pseudo-randomised sample of imprisoned and non-imprisoned subjects given the set of all confounders $\boldsymbol{L}$. We match each imprisoned subject on one or more non-imprisoned subjects based on the specific propensity score values.

A matching procedure using multiple mathcing groups was implemented to better match the results and therefore reduce the bias in the final estimates. The idea of matching within a defined group derived from a matching technique, where multiple control groups are matched to one exposure group (Stuart and Rubin, 2008). Hence, we joined sites into groups on which the matching was implemented. These groups were created, taking into account geographical proximity and epidemiological similarities, including the HIV

prevalence of a specific region. Over the whole dataset 7 of these groups were defined (group 1 - UK-EWnI, UK-S; group 2 - RU-IN, RU-StP, RU-V, RU-5s; group 3 - GR-A; group 4 - CZ, HU, PL-G; PL-GK, PL-W, PL-Ms; group 5 - EE-T, LV-5s, LV-R; group 6 – LU, NL-A, FI-7; group 7 - PT-P, SP-C, SP-MBS). Therefore, each imprisoned subject could only be matched to non-imprisoned subjects from the same defined group. The idea of matching being executed on a city-year level was also considered but abandoned on the account of scarcity of imprisoned subjects in some sites.

Within each group, the nearest neighbour $k : 1$ matching using a variable ratio matching algorithm (Stuart, 2010) with the propensity score difference was implemented. A limit to the maximum number of matches of $k = 4$ was set due to there being a large number of close matches, many of which were of the $4 : 1$ by nature, meaning that each imprisoned subject matched to the maximum number of 4 non-imprisoned subjects. The variable ratio matching algorithm induces flexibility of the matching ratio that aided in the groups where close matches were not so common, contributing to smaller bias in the final estimate. The variable ratio matching algorithm permitted imprisoned subjects to match to a maximum of four non-imprisoned subjects with the lowest respective propensity score differences. A tolerance measure equal to 10^{-5} was used so that the propensity score differences less than this measure were considered equal. In the case where more than four non-imprisoned subjects had a score difference of less than the tolerance measure, amongst those, four subjects were chosen randomly. However, where less or equal to four non-imprisoned subjects matched within the measure of tolerance, the matching resulted in these specified matches. This ensured that the proportion of non-imprisoned subjects would not be too dissimilar between the matching groups. In addition, matching using a variable ratio algorithm avoided the loss of several matches within the measure of tolerance, where lesser matches arose, improving further the overall quality of matching. We did not define the smallest accepted distance between the propensity scores of two subjects (caliper size) as a large number of close matches were identified among all groups.

The matching procedure was implemented with replacement, which allowed the same non-imprisoned subject to be matched to multiple imprisoned subjects. In addition, the variable ratio matching procedure allowed different number of non-imprisoned subjects to be matched to each imprisoned subject. Therefore, weights proportional to the number of non-imprisoned matched to each imprisoned subject were assigned. These weights were later accounted for in the analysis through the use of weighted regression. The propensity score matching was performed with R (version 3.6.1) using the function *Match* together with *MatchBalance* from the package *Matching* (Sekhon, 2011). Following this, the matched data sets from each group were then combined.

The goodness of matching was evaluated using the most common numerical balance measures of mean standardised differences and variance ratios between between the imprisoned and non-imprisoned for each covariate as given in equation 2 and equation 3 respectively. After propensity score matching, 97.4% of all absolute standardised mean differences were below the chosen threshold of 25%. Furthermore, 76.6% of these differences were below the even lower threshold of 10%. In addition, 93.4% of the variance ratios stayed within the interval from 0 to 2 . The large majority of below threshold differences and within boundary variance ratios suggested a relatively good balance between the two exposure groups. Additionally, to illustrate the matching effect, distributional balance plots for all first order variables were plotted (see Appendix 2).

3.3 Modelling

To model the effect of ever being imprisoned on the probability of having HIV among people who inject drugs, a generalised linear mixed model was estimated. The model contained all variables given in Table 1.1 with city-year accounting for the City and Year of data collection and the exception of Country, Recently sharing syringes and the Categorical number of sex partners as similar information acquired already using either variables of Ever

sharing syringes or Number of sex partners. We used a logistic model with logit link, where variables city-year and the matching group were considered as random effects. The weights received from the variable ratio matching algorithm were directly introduced into the regression model through weighted least squares. Due to the number of random effects being larger than one, the estimation of the model's log-likelihood and therefore the model parameters was performed using the Laplace approximation of the integrand. The final model is given in equation 10.

$$\ln\left[\frac{P\left(\text{"HIV"}=1|\text{group, city-year}\right)}{1-P\left(\text{"HIV"}=1|\text{group, city-year}\right)}\right]=\gamma_0+\boldsymbol{\gamma}\boldsymbol{L_2}+\alpha_{\text{group}}+\alpha_{\text{city-year}}.^{3} \tag{10}$$

As the imputation procedure along with matching and modelling was done 25 times, the final model estimates were pooled (using equations 4 and 5) to obtain single estimates for the model parameters along with the corresponding standard errors and p-values that should control for the overall type I error. The final estimates given as adjusted odds ratios with the 95% confidence interval for all model characteristics are presented in Table 3.1.

The model estimates show the adjusted odds of HIV positivity to be 32% higher among PWID who have ever been imprisoned compared to those never imprisoned (AOR 1.32, 95% CI 1.09 − 1.59). That is, ever imprisonment having a 29% increase in HIV risk among PWID (RR 1.29, 95% CI 1.07-1.51; AOR 1.32, 95% CI 1.09-1.59; HIV prevalence in never imprisoned 7.28% 95% CI 6.90-7.66%). Further, longer duration of drug injection (AOR 1.31, 95% CI 1.16 − 1.48), ever having shared needles or syringes (AOR 1.91, 95% CI 1.59−2.28) and a national higher wealth inequality measure (AOR 1.34, 95% CI 1.18−1.51) all have a positive effect on the probability of HIV positivity. In addition, the most recent main drug injected being Cocaine (AOR 2.70, 95%

[3]Here γ_0 is the intercept, $\boldsymbol{\gamma}$ is the parameter vector, $\boldsymbol{L_2}$ the set of individual level confounders and α_{group} and $\alpha_{\text{city-year}}$ the matching group and city-year specific random effects from the $N(0,\sigma^2_{group})$ and $N(0,\sigma^2_{city-year})$ distributions respectively, with σ^2_{group} and $\sigma^2_{city-year}$ denoting the variances of the random intercepts of group and city-year.

Table 3.1: Study characteristics with estimations on the effect of HIV presented in adjusted odds ratios. For categorical variables the number of HIV_+ and total subjects is given, for numerical variables the mean, range and standard deviation is given.

	HIV_+/total	**AOR**	**95% CI**
Socio-demographic characteristics			
Age (mean, range, SD) all	32.77, 13-78, 8.42	0.84	0.76-0.94
Age (mean, range, SD) HIV_+	33.61, 17-64, 7.61		
Sex			
Male	2378/32,251	0.77	0.65-0.91
Female	710/11,165	1	
Drug use characteristics			
Duration of injecting (mean, range, SD) all	11.19, 0.02-53, 8.04	1.31	1.16-1.48
Duration of injecting (mean, range, SD) HIV_+	14.02, 0.33-40, 7.44		
Frequency of injecting			
Less than daily	1701/17,551	0.90	0.76-1.07
Daily or more	1016/12,449	1	
Main drug injected			
Amphetamine	171/1597	1	
Cocaine	230/2030	2.70	1.73-4.22
Opioid	1796/22,143	1.52	1.05-2.18
Opioid & Cocaine	264/2282	2.16	1.33-3.53
Other	83/1304	1.55	0.88-2.72
Overdose			
yes	1039/2432	1.21	0.97-1.51
no	681/2928	1	
Ever sharing syringes			
yes	1697/11,619	1.91	1.59-2.28
no	1119/19,702	1	
Sexual behaviour			
Number of partners (mean, range, SD) all	3.36, 0-2400, 25.38	1.03	0.996-1.06
Number of partners (mean, range, SD) HIV_+	7.85, 0-2400, 72.92		
Environmental factors			
Opioid substitution therapy			
yes	1075/27,047	1.22	0.96-1.56
no	632/8915	1	
Main source of clean syringes			
NSP+outreach	773/2451	1	
Pharmacy	745/2368	0.72	0.59-0.88
Other	91/559	0.88	0.59-1.31
Ever in prison			
yes	1803/24,857	1.32	1.09-1.59
no	1239/17,491	1	
Study level measures			
Gini index (mean, range, SD)	33.54, 25.4-44, 4.76	1.34	1.18-1.51

CI $1.73 - 4.22$), Opioid (AOR 1.52, 95% CI $1.05 - 2.18$) or a combination of the two (AOR 2.16, 95% CI $1.33 - 3.53$) as compared to those whose primary most recent injection being amphetamines also have a positive effect on the probability of HIV positivity.

On the other hand, both socio-demographic characteristics considered in our analysis, including the increase in the subject's age (AOR 0.84, 95% CI $0.76 - 0.94$) or the subject being male (AOR 0.77, 95% CI $0.65 - 0.91$) had a protective effect on the HIV positivity among PWID. Similarly, having the primary source of syringes from the pharmacy (AOR 0.72, 95% CI $0.59 - 0.88$) compared to an outreach or a needle and syringe program also has a protective effect against HIV positivity among PWID.

3.4 Sensitivity analysis

To address the suitability of the propensity score technique and test the robustness of the estimation of ever imprisonment on HIV among people who inject drugs, a sensitivity analysis was executed. In this section, we conducted additional analysis on the imprisonment effect on HIV using multiple different approaches. The first sensitivity analysis follows straight from the propensity score matching, where the exposure effect is assessed within subsets defined by the estimated propensity scores. For the second sensitivity analysis, we estimate the effect on the raw imputed data, without any implementation of propensity score methods. Finally, we study the unadjusted effect of ever imprisonment on HIV both on the propensity score matched data and the unmatched imputed data by univariable regression.

3.4.1 Propensity score stratification

Propensity score stratification is a bias reduction technique in which subjects are categorised into strata according to their estimated propensity scores (Austin, 2011). As this stratification is based on the propensity score, being a balancing score, the distribution of the observed covariates should be directly comparable within each stratum. This allows estimation of the exposure effect to take place within strata, using regular modelling, after which the stratum-specific estimates can be combined into the overall exposure effect. Usually, the stratification is based on quintiles of the estimated propensity scores and the subjects are divided into five subgroups of equal size (Cochran, 1968). Defining the subgroups based on the propensity score quintiles means that higher ranks of subgroups come with increasing probabilities of ever being imprisoned, while by the definition of the propensity score, higher propensity score values indicate higher probabilities of ever being imprisoned.

For the stratum-specific estimation of imprisonment on having HIV among PWID, a generalised linear mixed model was used. The shape of the model was identical to that used for the propensity score matched data, given in equation 10, with the exception of random effects. The random effects for the

current model were Country and city-year. The random effect parameters here were altered from those used in the propensity score matching model, as here the defining of matching groups was not needed and to receive the highest precision in the random effects with the model being convergent. Due to no matching conducted in this analysis, the matching weights were discarded from the model. Similarly to the estimates for the matched data, the stratification estimates were pooled across the 25 unique imputed data sets. This resulted in the stratum-specific estimates, given as adjusted odds ratios with the 95% confidence intervals, of the imprisonment effect on HIV, shown in Table 3.2.

The fractional values in the subject count come from averaging the total number of subjects and ever imprisoned subjects over the 25 datasets. The combined exposure effect, shown in the same table, can be obtained by pooling the results from the stratum-specific estimates with the use of weights, defined by the proportion of subjects in a given stratum (Rudolph et al., 2016). For the assessment of the effectiveness of the stratification process, distributional balance plots for all variables were plotted by cohort (see Appendix 3).

Table 3.2: Stratum-specific estimation of imprisonment effect on HIV

Stratum	**Imprisoned**	**Subjects**	**AOR**	**95% CI**
1	2353.6	8762	1.78	1.47-2.16
2	4305.7	8761	1.63	1.30-2.04
3	5416.6	8761	1.30	1.02-1.65
4	6320.6	8761	1.15	0.83-1.58
5	7355.4	8762	0.94	0.68-1.29
Combined			1.32	1.17-1.49

The overall estimation of the imprisonment effect on HIV among PWID, given as an adjusted odds ratio was 1.32 (95% CI $1.17 - 1.49$), which is fairly similar to the estimate acquired from propensity score matching 1.32 (95% CI $1.09 - 1.59$).

Table 3.2 presents the stratum specific estimates of the imprisonment effect on HIV status among PWID. For strata one to three a positive effect of HIV positivity is estimated among PWID who have ever been imprisoned compared to those never imprisoned (stratum 1 AOR 1.78, 95% CI 1.47 − 2.16), (stratum 2 AOR 1.63, 95% CI 1.30 − 2.04), (stratum 3 AOR 1.30, 95% CI 1.02 − 1.65). For strata four and five no statistically significant effect was established.

3.4.2 Estimation without propensity score methods

The sensitivity analysis also included a direct approach of estimating the probability of having HIV among PWID without the use of propensity score methods. This direct approach meant that a model was fitted immediately after the imputation steps with no matching taking place. Modelling directly the imputed data allowed us to compare the two approaches and to understand the impact of propensity score matching on the estimates within our data set.

A generalised linear mixed model, analogous to that used for the propensity score matched data, given in equation 10 was used. The difference in the models was in the random effect parameters, where Country and city-year were used as random effects. The random effect parameters here were altered from those used in the propensity score matching model, as here the defining of matching groups was not needed and to receive the highest precision in the random effects with the model being convergent. As no matching had occurred, the matching weights were also discarded from the model. Similarly with the other approaches, the estimations were pooled across the 25 imputed data sets. The final estimates are given in Table 3.3 together with the estimates of the propensity score matched model for comparison.

The estimate of ever being in prison for the unmatched model 1.48 (95% CI 1.34 − 1.63) differs somewhat from the estimate in the propensity score matched model 1.32 (95% CI 1.09−1.59). Comparing the confidence intervals of the two estimates, we see overlapping, however, the unmatched model

Table 3.3: Effect estimates of having HIV for both propensity score matched data and unmatched data presented in adjusted odds ratios for all study variables

	Multivariable propensity score matched model		**Multivariable unmatched model**	
	AOR	**95% CI**	**AOR**	**95% CI**
Socio-demographic characteristics				
Age	0.84	0.76-0.94	0.89	0.82-0.96
Sex				
Male	0.77	0.65-0.91	0.88	0.79-0.97
Female	1		1	
Drug use characteristics				
Duration of injecting	1.31	1.16-1.48	1.33	1.23-1.43
Frequency of injecting				
Less than daily	0.90	0.76-1.07	0.83	0.73-0.93
Daily or more	1		1	
Main drug injected				
Amphetamine	1		1	
Cocaine	2.70	1.73-4.22	2.62	1.90-3.61
Opioid	1.52	1.05-2.18	1.54	1.23-1.92
Opioid & Cocaine	2.16	1.33-3.53	2.50	1.76-3.54
Other	1.55	0.88-2.72	1.60	1.19-2.14
Overdose				
yes	1.21	0.97-1.51	1.31	1.07-1.59
no	1		1	
Ever sharing syringes				
yes	1.91	1.59-2.28	1.68	1.51-1.86
no	1		1	
Sexual behaviour				
Number of partners	1.03	0.996-1.06	1.03	1.01-1.06
Environmental factors				
Opioid substitution therapy				
yes	1.22	0.96-1.56	1.25	1.05-1.49
no	1		1	
Main source of clean syringes				
NSP+outreach	1		1	
Pharmacy	0.72	0.59-0.88	0.74	0.62-0.88
Other	0.88	0.59-1.31	0.89	0.61-1.28
Ever in prison				
yes	1.32	1.09-1.59	1.48	1.34-1.63
no	1		1	
Study level measures				
Gini index	1.34	1.18-1.51	1.25	1.10-1.44

gives a seemingly larger effect to ever imprisonment on acquiring HIV among PWID.

Table 3.3 also indicates relatively similar effect estimates for both of the models, as the direction of the adjusted odds ratios are equivalent for all parameters included in both models. Additionally, all significant effect estimates for the propensity score matched model are also statistically significant for the unmatched model. Nevertheless, for the unmatched model, Frequency of injecting of less than daily compared to daily or more 0.83 (95% CI $0.73 - 0.93$), Other main drug injected compared to Amphetamine 1.60 (95% CI $1.19 - 2.14$), ever experiencing overdose 1.31 (95% CI $1.07 - 1.59$), higher number of sex partners 1.03 (95% CI $1.01 - 1.06$) and ever receiving opioid substitution therapy 1.25 (95% CI $1.05 - 1.49$) are all statistically significant in contrast to the results of the propensity score matched model.

Furthermore, the point estimates of the propensity score matched model for Male gender, Ever sharing syringes and Ever in prison are not covered by the confidence intervals of the unmatched model. However, when comparing the corresponding confidence intervals of the two models, we see coinciding of all respective intervals.

3.4.3 Unadjusted estimations on imputed and matched data

Lastly, we examine the unadjusted effect of imprisonment on HIV among PWID both on the propensity score matched data and the unmatched imputed data by univariable regression. For the propensity score matched data we used a generalised linear mixed model, given in equation 10, where the parameter vector $\boldsymbol{L_2}$ contained only the effect of imprisonment. For the unmatched imputed data, the same mixed model was used with the difference in the matching group specific random effect, which was substituted for the Country specific random effect, as no matching had taken place in the imputed data set. Similarly to the other methods, the estimations were pooled across the 25 data sets to receive the final pooled estimates of imprisonment on HIV for both approaches.

The unadjusted effect estimate on the raw imputed data of 1.76 (95% CI $1.61 - 1.94$) alters considerably from the unajdusted effect estimate on the propensity score matched data of 1.27 (95% CI $1.05 - 1.53$), both given as odds ratios, which is noticeably closer to the adjusted estimate of the propensity score matched model 1.32 (95% CI $1.09 - 1.59$). Although the two unadjusted estimates share the direction, the unmatched effect is evidently overestimated, yielding confidence intervals with no overlap. This result supports the findings of the propensity score matching procedure, where a relatively good balance between the two imprisonment groups was identified. Consequently, this demonstrates that in the propensity score matched data the confounding between the confounder set and imprisonment has been largely adjusted for, with the confounding relationship having been effectively erased.

4 Discussion

In general, traditional regression analysis is appropriate in estimating causal inference. The causal effect estimates of an exposure on an outcome result from conditioning on the confounder set (see example diagram in Figure 2.1), hence blocking all causal paths intercepting it. However, as the confounder set is often high-dimensional, statistical inference methods exclusively relying on regression analysis rarely work in practice due to misspecification of the regression model and the issues with the curse of dimensionality, where adequate sample size is seldom achieved. While traditional regression analysis focuses on modelling the association between the confounders and the outcome, any misspecification to the model lead to biased effect estimators and therefore inferring causal effects becomes inadequate. (Hernán and Robins, 2020)

The problem of misspecification is particularly significant when little overlap between the exposure groups exists (Rubin, 1997). In this case, extrapolation is needed in order to infer causal effects. Propensity score methods, on the other hand, are not susceptible to extrapolation or the curse of dimensionality as they make use of modelling the association between the exposure and confounders (Vansteelandt and Daniel, 2014). This, however, leads to larger standard errors when compared against regression methods, where the uncertainty in extrapolation is ignored, and therefore wider confidence intervals, also seen in the results of this study (see Table 3.3).

In addition, the study revealed an exposure response type effect in the propensity score stratification sensitivity analysis. Table 3.2 presented the stratum specific estimates of the imprisonment effect on HIV status among PWID. Interestingly, the decrease of point-wise estimates together with the 95% confidence intervals with higher ranks of subgroups suggest a diminishing effect of imprisonment on HIV among PWID. Furthermore, in the bottom two subgroups where the probability of imprisonment was the highest, there is no longer a statistically significant difference between the ever imprisoned and non-imprisoned with regard to HIV positivity among PWID. Hence, as the

subgroups were defined based on the quintiles of the estimated propensity scores, our results indicate that for subgroups, where the probability of ever being imprisoned is higher, the adjusted odds ratio of HIV positivity were lower among ever imprisoned PWID compared to those never imprisoned.

This study has also important caveats that should be noted. Foremost, the analysis data set being not a complete random sample, causal inference estimations will always be subject to dispute, leading from the assumption of no unmeasured confounding where one can never affirm it to be completely satisfied. As this study tackles the stigmatised and hidden issues of imprisonment and HIV, both of which challenging for the subjects and a population hard to reach by the investigators, a completely random sample would presently be unfeasible. Specifically, randomisation at an individual level not even possible. The sampling methods in this study were also dissimilar across sites, varying between convenience sampling, respondent driven sampling, snowball sampling and purposive sampling methods, all adding to further divergence of the collected data. In addition, the unification of such extensive base data collected in different years will always lead to a loss of information in some regard. Nevertheless, aim of minimal data loss was pursued during the whole study process.

Estimating effect sizes using propensity score techniques could also pose some problems and the assumptions for using propensity scores should be carefully examined when using this method. For example, the propensity score should be constructed on the entire confounder set $\boldsymbol{L}$, an assumption that is difficult to validate in practice. In our analysis, we have included as many variables as possible and as literature has suggested, including many higher order terms, and thus we expect the confounder set $\boldsymbol{L}$ to be well described with our choice although existence of additional variables cannot be ruled out.

In literature, propensity score techniques have also been greeted with some reluctance. King and Nielsen (King and Nielsen, 2019) propose two main critiques to the propensity score matching method. Firstly, that matching on the propensity score attempts to imitate a completely randomised design instead of a more efficient randomised block design. Correct inference

is therefore yet more reliant on the propensity score model being correctly specified, as covariate balance is attained evenly from all covariates included into the propensity score model. An alternative would be matching on the Mahalanobis distance, where prior knowledge of variable importance can be taken into account. Secondly, that propensity score matching may give rise to the *"propensity score matching paradox"*. The paradox states that data sets with good underlying covariate balance may lead to increased imbalance in the covariate distributions between exposure groups after the matching procedure caused by too narrow of a chosen caliper size. As in the current study, all exposed subjects had an adequate number of matches, many of which more than one, we believe that in this study this issue was under control.

Conclusion

The thesis explored the causal effect of ever imprisonment on Human Immunodeficiency Virus (HIV) among people who inject drugs (PWID) from individual patient data collected in the project *"European Study Group for Mathematical Modelling and Epidemiological Analysis of Drug Related Infectious Diseases"*, coordinated by the *European Monitoring Centre for Drugs and Drug Addiction*. The thesis is a comprehensive overview of the article (Wiessing, Uusküla, and Rannap et al., 2021) submitted to *The Addiction*. We introduced propensity score methods as bias reduction techniques, applied to the unified HIV data set, to support in the estimating of this causal effect. To the knowledge of the authors, this study is the largest individual patient data meta-analysis conducted as yet, assessing the aforementioned causal relationship.

Our results suggest that imprisonment significantly increases risk for HIV infection among PWID (RR 1.29, 95% CI 1.07-1.51; AOR 1.32, 95% CI 1.09-1.59; HIV prevalence in never imprisoned 7.28% 95% CI 6.90-7.66%). A recent systematic review and meta-analysis (Stone et al., 2018) proposed ever imprisonment having a comparable 25% increase in HIV risk among PWID, supporting our main estimate. Besides the effect of imprisonment, our analysis highlights several other factors to be associated with the risk of HIV, including the age and sex of the subjects, injection duration, main drug injected, ever sharing syringes, main source of clean syringes at the individual level, and Gini index, as a societal level factor (indicator of socio-economic inequality).

Although the results of our study are concordant with previous research, we need to acknowledge the plausible deviation from the assumption of conditional exchangeability and therefore from the assumption of no unmeasured confounding in our estimations of the propensity score. Nevertheless, due to the construction of our propensity score model and the sensitivity analysis results, we argue that this assumption is highly likely to be valid. Furthermore, the assumption of no unmeasured confounding is also assumed to hold for all

other regression-based approaches that estimate the treatment effect in observational studies (Austin, 2011). Consequently, due to the large amount of sites with different definitions of collected variables and the variability in the specific recall periods, the data unification process will always lead to lower accuracy of study data compared to the site-specific data. To mitigate this, the unification process was executed so to minimise the potential bias leading from the unification process. In light of these preceding limitations, further research is needed to assess the constraints resulting from the propensity score methods and data unification issues and to more precisely describe the association of HIV transmission with regard to imprisonment among PWID. While these biases may influence the estimate of imprisonment-HIV association, they seem unlikely to have caused the clear patterns observed in this study.

In conclusion, the study confirms a risk-increasing causal relationship between HIV risk and the history of incarceration among community-recruited PWID in Europe. The methods of propensity scores, including matching and stratification, both yield results similar to previous research. We believe that the implementation of the propensity score methods help in the accuracy of the causal estimation with regard to reducing the bias naturally present in observational studies.

References

Agresti, Alan (2013). *Categorical data analysis*. 3rd ed. Wiley series in probability and statistics 792. Hoboken, NJ: Wiley. ISBN: 9780470463635.

Austin, Peter C. (May 2011). "An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies". en. In: *Multivariate Behavioral Research* 46.3, pp. 399–424. ISSN: 0027-3171, 1532-7906. DOI: 10.1080/00273171.2011.568786. URL: http://www.tandfonline.com/doi/abs/10.1080/00273171.2011.568786 (visited on 03/05/2021).

Azur, Melissa J., Elizabeth A. Stuart, Constantine Frangakis, and Philip J. Leaf (Mar. 2011). "Multiple imputation by chained equations: what is it and how does it work?: Multiple imputation by chained equations". en. In: *International Journal of Methods in Psychiatric Research* 20.1, pp. 40–49. ISSN: 10498931. DOI: 10.1002/mpr.329. URL: http://doi.wiley.com/10.1002/mpr.329 (visited on 03/07/2021).

Breslow, N. E. and D. G. Clayton (Mar. 1993). "Approximate Inference in Generalized Linear Mixed Models". In: *Journal of the American Statistical Association* 88.421, p. 9. ISSN: 01621459. DOI: 10.2307/2290687. URL: https://www.jstor.org/stable/2290687?origin=crossref (visited on 03/13/2021).

Buuren, Stef van (Mar. 2012). *Flexible Imputation of Missing Data*. en. 0th ed. Chapman and Hall/CRC. ISBN: 9780429065408. DOI: 10.1201/b11826. URL: https://www.taylorfrancis.com/books/9781439868256 (visited on 03/29/2021).

Buuren, Stef van and Karin Groothuis-Oudshoorn (2011). "**mice** : Multivariate Imputation by Chained Equations in *R*". en. In: *Journal of Statistical Software* 45.3. ISSN: 1548-7660. DOI: 10.18637/jss.v045.i03. URL: http://www.jstatsoft.org/v45/i03/ (visited on 03/07/2021).

Cochran, W. G. (June 1968). "The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies". In: *Biometrics* 24.2, p. 295. ISSN: 0006341X. DOI: 10.2307/2528036. URL: https://www.jstor.org/stable/2528036?origin=crossref (visited on 09/20/2021).

Degenhardt, Louisa, Amy Peacock, Samantha Colledge, Janni Leung, Jason Grebely, Peter Vickerman, Jack Stone, Evan B Cunningham, Adam Trickey, Kostyantyn Dumchev, Michael Lynskey, Paul Griffiths, Richard P Mattick, Matthew Hickman, and Sarah Larney (Dec. 2017). "Global prevalence of injecting drug use and sociodemographic characteristics and prevalence of HIV, HBV, and HCV in people who inject drugs: a multistage systematic review". en. In: *The Lancet Global Health* 5.12, e1192–e1207. ISSN: 2214109X. DOI: 10.1016/S2214-109X(17)30375-3. URL: https://linkinghub.elsevier.com/retrieve/pii/S2214109X17303753 (visited on 11/13/2021).

Dehejia, Rajeev H. and Sadek Wahba (Dec. 1999). "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs". en. In: *Journal of the American Statistical Association* 94.448, pp. 1053–1062. ISSN: 0162-1459, 1537-274X. DOI: 10.1080/01621459.1999.10473858. URL: http://www.tandfonline.com/doi/abs/10.1080/01621459.1999.10473858 (visited on 03/01/2021).

Grant, R. L. (Jan. 2014). "Converting an odds ratio to a range of plausible relative risks for better communication of research findings". en. In: *BMJ* 348.jan24 1, f7450–f7450. ISSN: 1756-1833. DOI: 10.1136/bmj.f7450. URL: https://www.bmj.com/lookup/doi/10.1136/bmj.f7450 (visited on 11/13/2021).

Greene, William H. (2003). *Econometric analysis*. 5th ed. Upper Saddle River, N.J: Prentice Hall. ISBN: 9780130661890.

Haneuse, Sebastien (Apr. 2016). "Distinguishing Selection Bias and Confounding Bias in Comparative Effectiveness Research". en. In: *Medical Care* 54.4, e23–e29. ISSN: 0025-7079. DOI: 10.1097/MLR.0000000000000011. URL: https://journals.lww.com/00005650-201604000-00013 (visited on 03/20/2021).

Hernán, Miguel and James M. Robins (2020). *Causal inference*. eng. Chapman & Hall/CRC monographs on statistics & applied probability. OCLC: 1009571686. ISBN: 9781420076165.

Holland, Paul W. (Dec. 1986). "Statistics and Causal Inference". en. In: *Journal of the American Statistical Association* 81.396, pp. 945–960. ISSN: 0162-

1459, 1537-274X. DOI: 10.1080/01621459.1986.10478354. URL: http://www.tandfonline.com/doi/abs/10.1080/01621459.1986.10478354 (visited on 03/20/2021).

Jiang, Jiming (2007). *Linear and generalized linear mixed models and their applications*. Springer series in statistics. OCLC: ocm77256604. New York ; London: Springer. ISBN: 9780387479415 9780387479460.

King, Gary and Richard Nielsen (Oct. 2019). "Why Propensity Scores Should Not Be Used for Matching". en. In: *Political Analysis* 27.4, pp. 435–454. ISSN: 1047-1987, 1476-4989. DOI: 10.1017/pan.2019.11. URL: https://www.cambridge.org/core/product/identifier/S1047198719000111/type/journal_article (visited on 11/25/2021).

Kivimets, Kristel, Anneli Uusküla, Jeffrey V. Lazarus, and Kristi Ott (Dec. 2018). "Hepatitis C seropositivity among newly incarcerated prisoners in Estonia: data analysis of electronic health records from 2014 to 2015". en. In: *BMC Infectious Diseases* 18.1, p. 339. ISSN: 1471-2334. DOI: 10.1186/s12879-018-3242-2. URL: https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-018-3242-2 (visited on 10/17/2021).

Lee, Brian K., Justin Lessler, and Elizabeth A. Stuart (Feb. 2010). "Improving propensity score weighting using machine learning". en. In: *Statistics in Medicine* 29.3, pp. 337–346. ISSN: 02776715. DOI: 10.1002/sim.3782. URL: http://doi.wiley.com/10.1002/sim.3782 (visited on 03/01/2021).

Little, Roderick J. A. (July 1988). "Missing-Data Adjustments in Large Surveys". In: *Journal of Business & Economic Statistics* 6.3, p. 287. ISSN: 07350015. DOI: 10.2307/1391878. URL: https://www.jstor.org/stable/1391878?origin=crossref (visited on 03/21/2021).

Raudenbush, Stephen W., Meng-Li Yang, and Matheos Yosef (Mar. 2000). "Maximum Likelihood for Generalized Linear Models with Nested Random Effects via High-Order, Multivariate Laplace Approximation". In: *Journal of Computational and Graphical Statistics* 9.1, p. 141. ISSN: 10618600. DOI: 10.2307/1390617. URL: https://www.jstor.org/stable/1390617?origin=crossref (visited on 03/14/2021).

Rosenbaum, Paul R. and Donald B. Rubin (1983). "The central role of the propensity score in observational studies for causal effects". en. In:

Biometrika 70.1, pp. 41–55. ISSN: 0006-3444, 1464-3510. DOI: 10.1093/biomet/70.1.41. URL: https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/70.1.41 (visited on 02/27/2021).

Rubin, Donald B., ed. (June 1987). *Multiple Imputation for Nonresponse in Surveys*. en. Wiley Series in Probability and Statistics. Hoboken, NJ, USA: John Wiley & Sons, Inc. ISBN: 9780470316696 9780471087052. DOI: 10.1002/9780470316696. URL: http://doi.wiley.com/10.1002/9780470316696 (visited on 03/07/2021).

– (Oct. 1997). "Estimating Causal Effects from Large Data Sets Using Propensity Scores". en. In: *Annals of Internal Medicine* 127.8_Part_2, p. 757. ISSN: 0003-4819. DOI: 10.7326/0003-4819-127-8_Part_2-199710151-00064. URL: http://annals.org/article.aspx?doi=10.7326/0003-4819-127-8_Part_2-199710151-00064 (visited on 11/19/2021).

Rubin, Donald B (2001). "Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation". In: *Matched Sampling for Causal Effects*, 365–382. DOI: 10.1017/cbo9780511810725.030.

Rubin, Donald B. and Nathaniel Schenker (June 1986). "Multiple Imputation for Interval Estimation from Simple Random Samples with Ignorable Nonresponse". en. In: *Journal of the American Statistical Association* 81.394, pp. 366–374. ISSN: 0162-1459, 1537-274X. DOI: 10.1080/01621459.1986.10478280. URL: http://www.tandfonline.com/doi/abs/10.1080/01621459.1986.10478280 (visited on 03/21/2021).

Rubin, Donald B. and Neal Thomas (Mar. 1996). "Matching Using Estimated Propensity Scores: Relating Theory to Practice". In: *Biometrics* 52.1, p. 249. ISSN: 0006341X. DOI: 10.2307/2533160. URL: https://www.jstor.org/stable/2533160?origin=crossref (visited on 03/05/2021).

Rudolph, Kara E., K. Ellicott Colson, Elizabeth A. Stuart, and Jennifer Ahern (Nov. 2016). "Optimally combining propensity score subclasses". en. In: *Statistics in Medicine* 35.27, pp. 4937–4947. ISSN: 0277-6715, 1097-0258. DOI: 10.1002/sim.7046. URL: https://onlinelibrary.wiley.com/doi/10.1002/sim.7046 (visited on 01/06/2022).

Sekhon, Jasjeet S. (2011). "Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The **Matching** Package for

R". en. In: *Journal of Statistical Software* 42.7. ISSN: 1548-7660. DOI: 10.18637/jss.v042.i07. URL: http://www.jstatsoft.org/v42/i07/ (visited on 02/27/2021).

Stewart, Lesley A. and Jayne F. Tierney (Mar. 2002). "To IPD or not to IPD?: Advantages and Disadvantages of Systematic Reviews Using Individual Patient Data". en. In: *Evaluation & the Health Professions* 25.1, pp. 76–97. ISSN: 0163-2787, 1552-3918. DOI: 10.1177/0163278702025001006. URL: http://journals.sagepub.com/doi/10.1177/0163278702025001006 (visited on 10/16/2021).

Stone, Jack, Hannah Fraser, Aaron G Lim, Josephine G Walker, Zoe Ward, Louis MacGregor, Adam Trickey, Sam Abbott, Steffanie A Strathdee, Daniela Abramovitz, Lisa Maher, Jenny Iversen, Julie Bruneau, Geng Zang, Richard S Garfein, Yung-Fen Yen, Tasnim Azim, Shruti H Mehta, Michael-John Milloy, Margaret E Hellard, Rachel Sacks-Davis, Paul M Dietze, Campbell Aitken, Malvina Aladashvili, Tengiz Tsertsvadze, Viktor Mravčík, Michel Alary, Elise Roy, Pavlo Smyrnov, Yana Sazonova, April M Young, Jennifer R Havens, Vivian D Hope, Monica Desai, Ellen Heinsbroek, Sharon J Hutchinson, Norah E Palmateer, Andrew McAuley, Lucy Platt, Natasha K Martin, Frederick L Altice, Matthew Hickman, and Peter Vickerman (Dec. 2018). "Incarceration history and risk of HIV and hepatitis C virus acquisition among people who inject drugs: a systematic review and meta-analysis". en. In: *The Lancet Infectious Diseases* 18.12, pp. 1397–1409. ISSN: 14733099. DOI: 10.1016/S1473-3099(18)30469-9. URL: https://linkinghub.elsevier.com/retrieve/pii/S1473309918304699 (visited on 10/17/2021).

Stuart, Elizabeth A. (Feb. 2010). "Matching Methods for Causal Inference: A Review and a Look Forward". en. In: *Statistical Science* 25.1, pp. 1–21. ISSN: 0883-4237. DOI: 10.1214/09-STS313. URL: http://projecteuclid.org/euclid.ss/1280841730 (visited on 02/27/2021).

Stuart, Elizabeth A. and Donald B. Rubin (Sept. 2008). "Matching With Multiple Control Groups With Adjustment for Group Differences". en. In: *Journal of Educational and Behavioral Statistics* 33.3, pp. 279–306. ISSN: 1076-9986, 1935-1054. DOI: 10.3102/1076998607306078. URL: http://

journals.sagepub.com/doi/10.3102/1076998607306078 (visited on 09/18/2021).

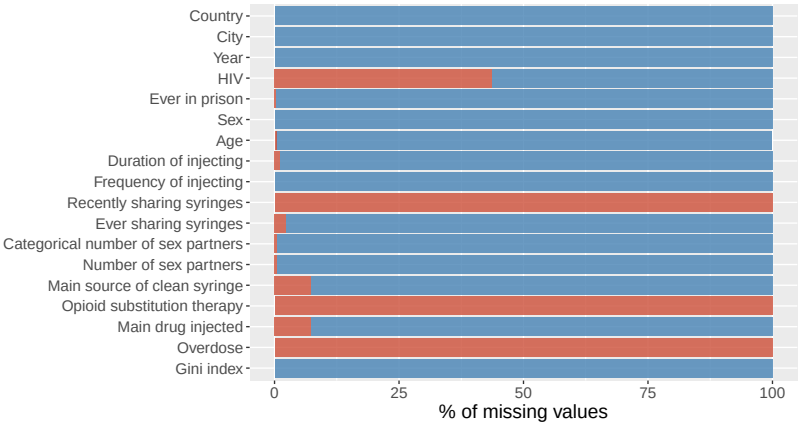
Vansteelandt, S. and R.M. Daniel (Oct. 2014). "On regression adjustment for the propensity score: S. VANSTEELANDT AND R. M. DANIEL". en. In: *Statistics in Medicine* 33.23, pp. 4053–4072. ISSN: 02776715. DOI: 10.1002/sim.6207. URL: https://onlinelibrary.wiley.com/doi/10.1002/sim.6207 (visited on 11/19/2021).

Weiss, Neil A. (2006). *A course in probability*. Pearson Addison Wesley.

Wiessing, Lucas, Anneli Uusküla, and Jürgen Rannap et al. (June 2021). "A history of incarceration is associated with HIV infection among community-recruited people who inject drugs in Europe - a propensity score matched individual level meta-analysis of cross-sectional studies". en. In: *The Addiction - submission*.

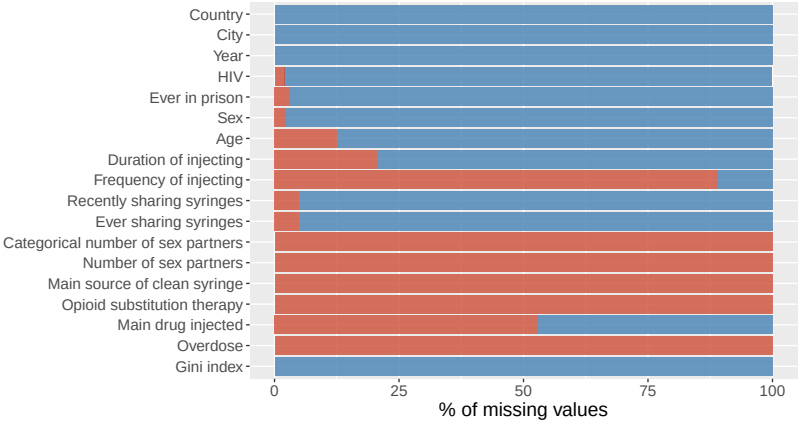
Zakrison, T. L., P. C. Austin, and V. A. McCredie (June 2018). "A systematic review of propensity score methods in the acute care surgery literature: avoiding the pitfalls and proposing a set of reporting guidelines". en. In: *European Journal of Trauma and Emergency Surgery* 44.3, pp. 385–395. ISSN: 1863-9933, 1863-9941. DOI: 10.1007/s00068-017-0786-6. URL: http://link.springer.com/10.1007/s00068-017-0786-6 (visited on 03/05/2021).

Appendix 1. Missingness patterns by country

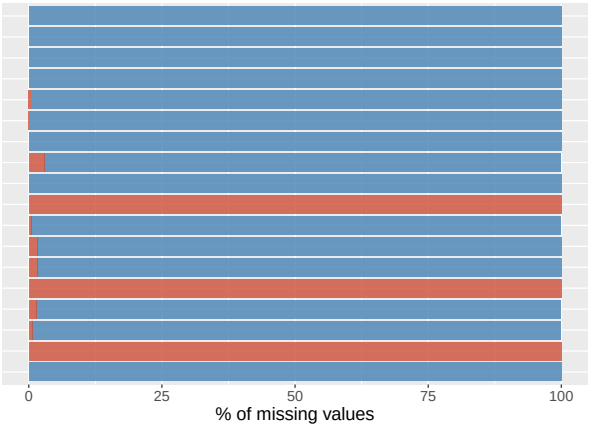


(a) Czech Republic

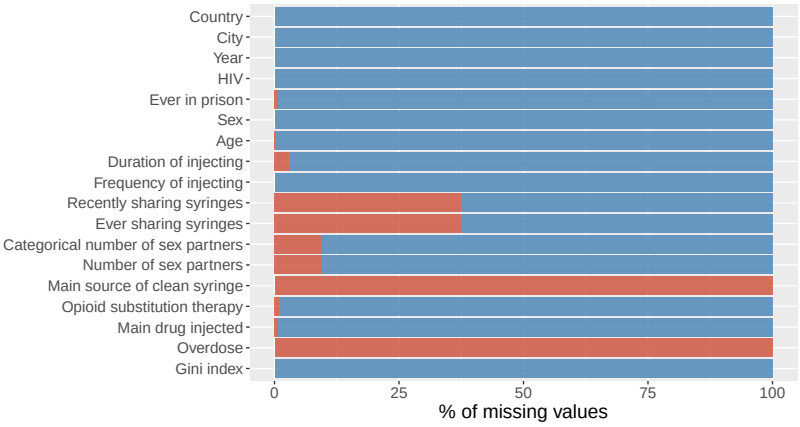
(b) Estonia



(c) Finland

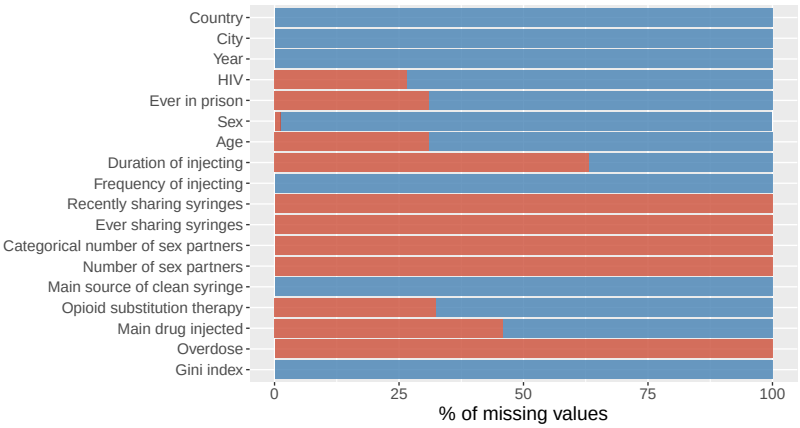


(d) Greece

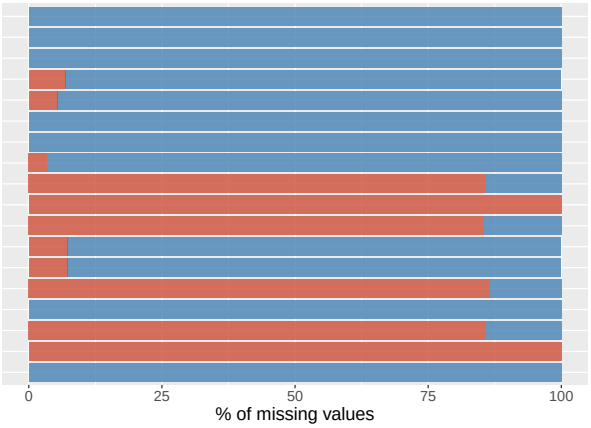


(e) Hungary

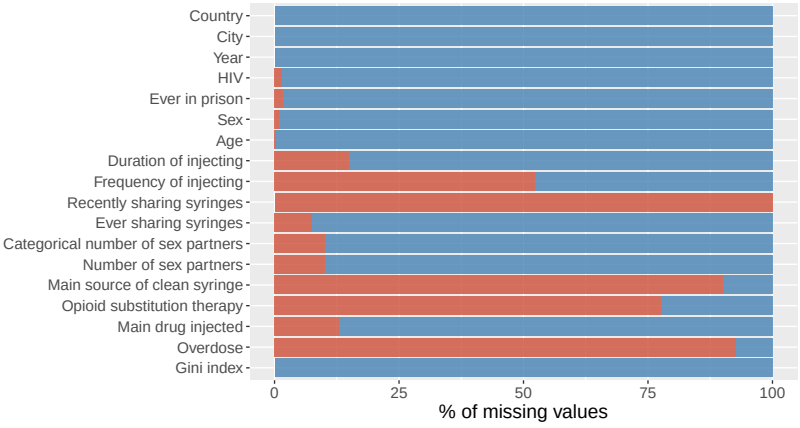
(f) Latvia



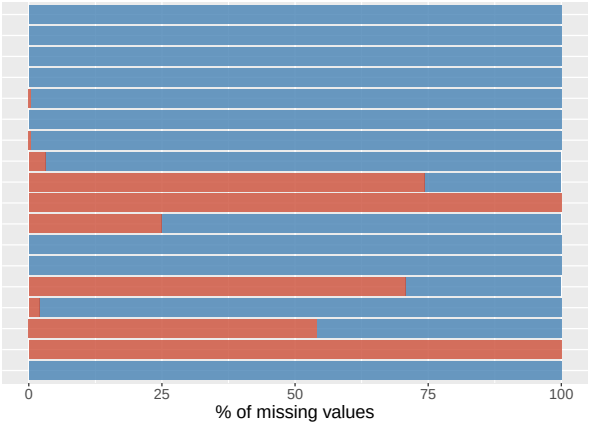
(g) Luxembourg



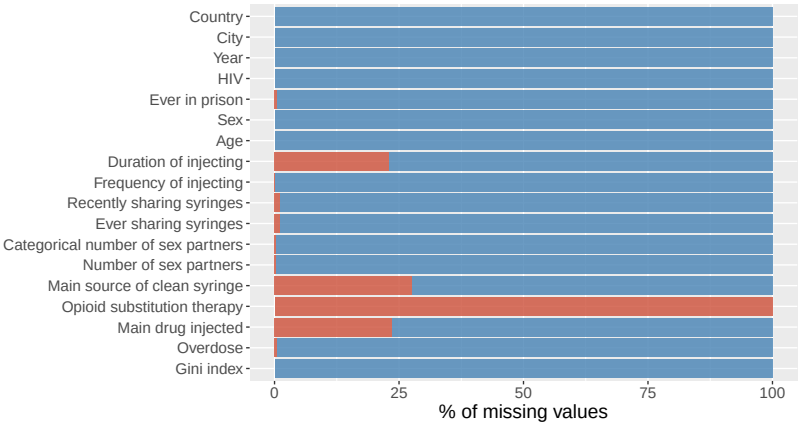
(h) Netherlands



(i) Poland

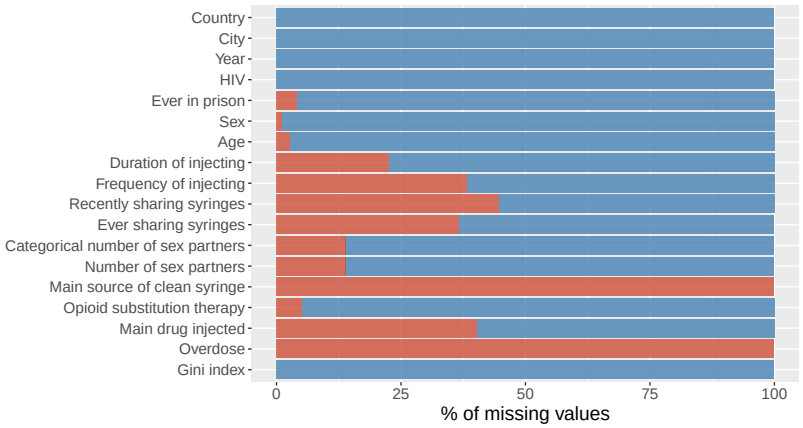


(j) Portugal



(k) Russia

(l) Spain



(m) United Kingdom

Figure A1: Proportion of missing (red) and observed (blue) data points by country

Appendix 2. Distributional balance plots of matched data

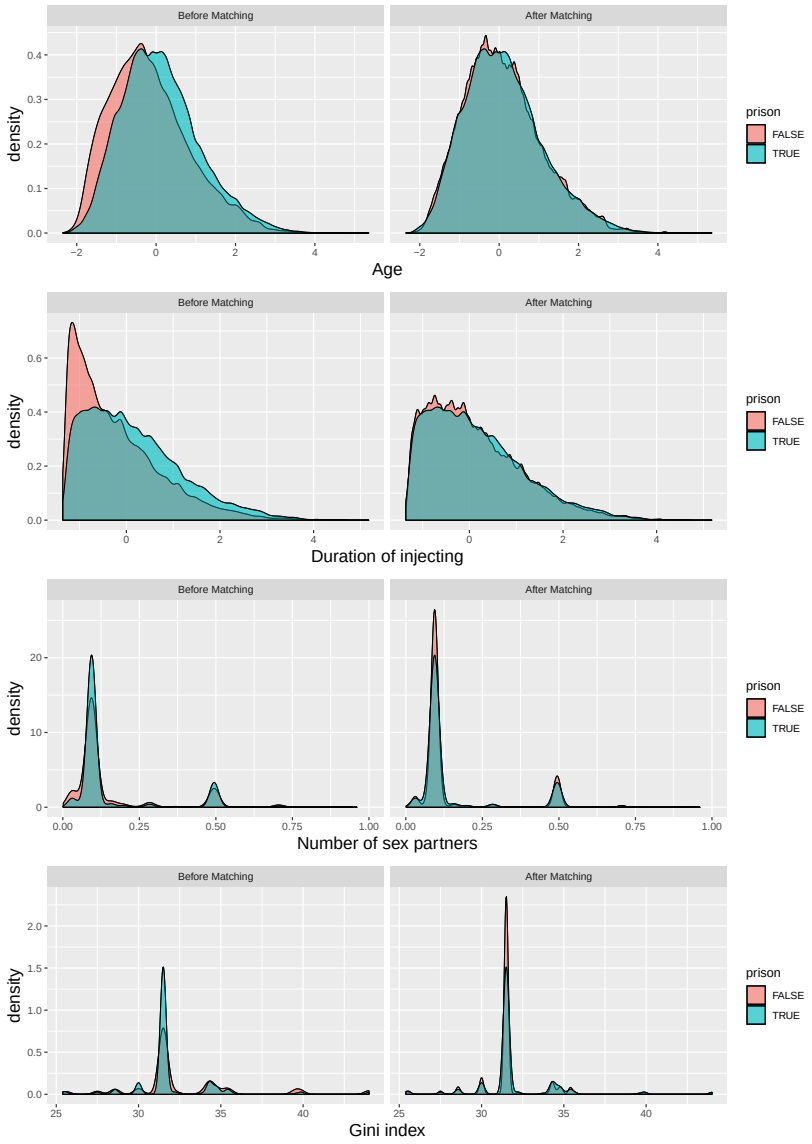


Figure A2: Density plots of numerical variables before and after matching by imprisonment


Before Matching
After Matching
ratio
1.00
0.75
0.50
0.25
0.00
female
male
Sex
FALSE
TRUE
HIV
Daily or more
Less than daily
Frequency of injecting
0
1
Ever sharing syringes
prison
FALSE
TRUE

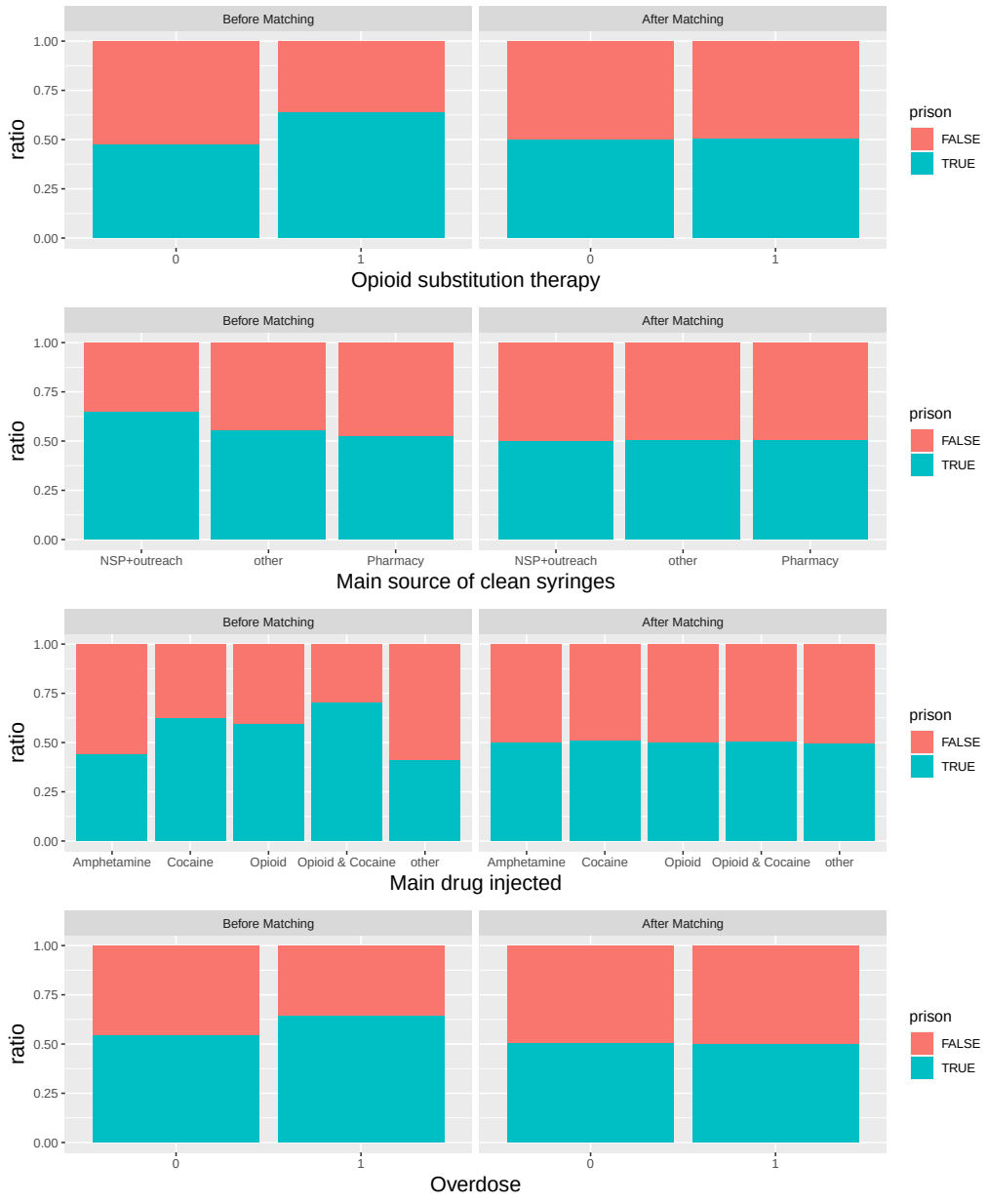


Figure A3: Proportionality plots of categorical variables before and after matching by imprisonment

Appendix 3. Distributional balance plots of stratified data

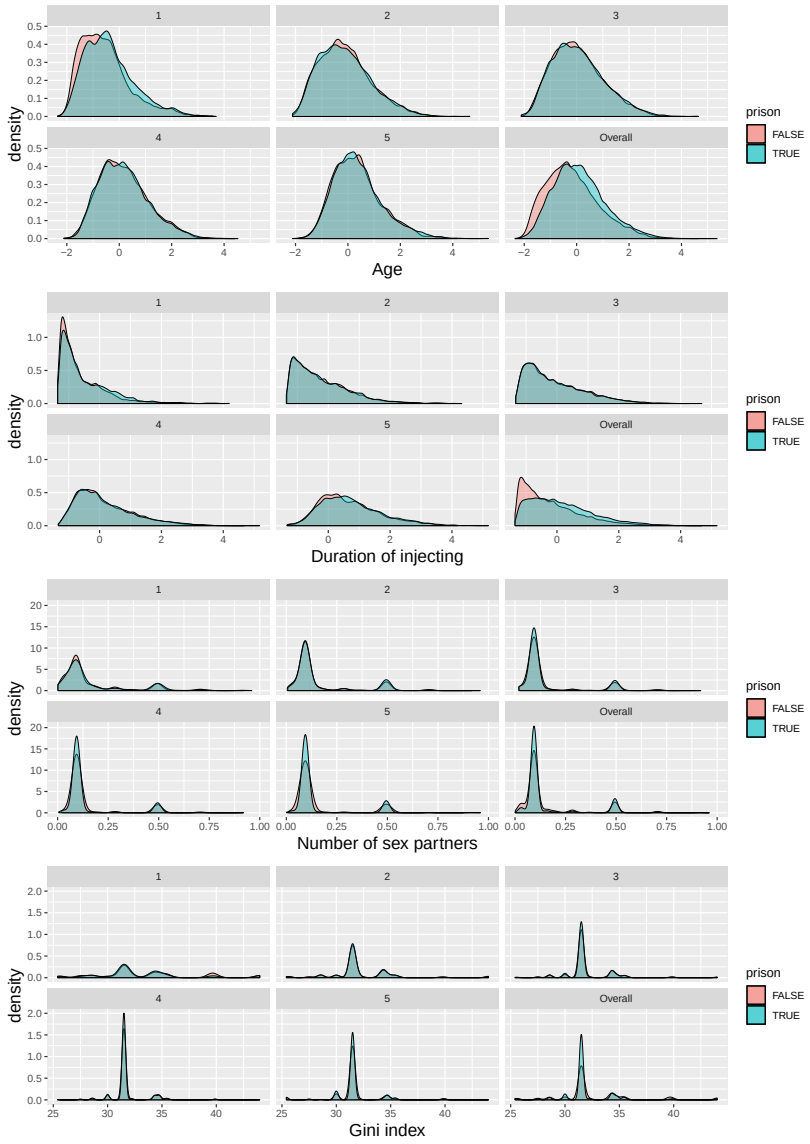
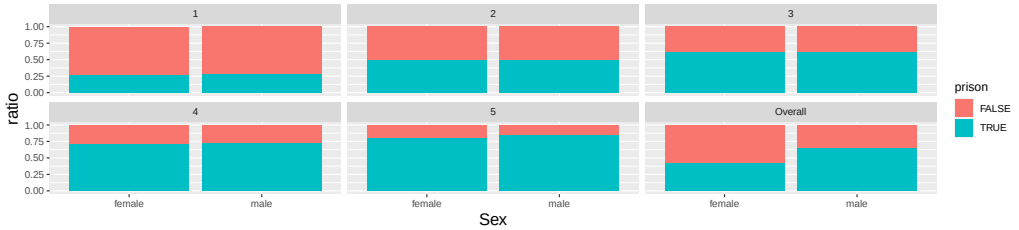
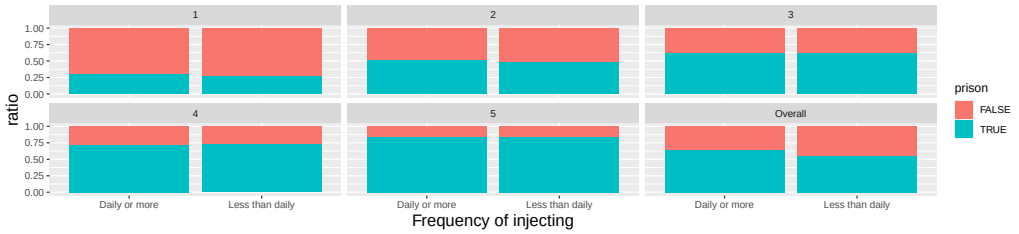


Figure A4: Density plots of numerical variables by cohort compared to overall by imprisonment


1
2
3
4
5
Overall
ratio
1.00
0.75
0.50
0.25
0.00
female
male
prison
FALSE
TRUE
Sex


1
2
3
4
5
Overall
ratio
1.00
0.75
0.50
0.25
0.00
FALSE
TRUE
prison
FALSE
TRUE
HIV


1
2
3
4
5
Overall
ratio
1.00
0.75
0.50
0.25
0.00
Daily or more
Less than daily
prison
FALSE
TRUE
Frequency of injecting

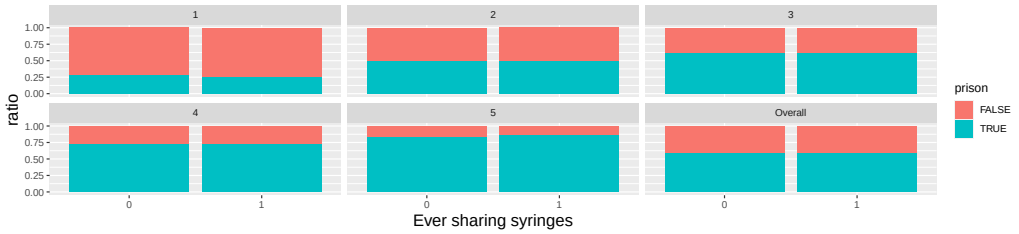

1
2
3
4
5
Overall
ratio
1.00
0.75
0.50
0.25
0.00
0
1
prison
FALSE
TRUE
Ever sharing syringes

Figure A5: Proportionality plots of categorical variables by cohort compared to overall by imprisonment

Non-exclusive licence to reproduce thesis and make thesis public

I, Jürgen Rannap,

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright, Causal effect of incarceration history on HIV susceptibility among people who inject drugs: an individual-level meta-analysis, supervised by Krista Fischer, Anneli Uusküla and Sven Erik Ojavee.

2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.

4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Jürgen Rannap
07/02/2022