

Discriminating Between Premigration and Postmigration HIV Acquisition Using Surveillance Data

Nikos Pantazis, PhD,^a Magdalena Rosinska, PhD,^b Ard van Sighem, PhD,^c Chantal Quinten, PhD,^d Teymur Noori, MSc,^d Fiona Burns, PhD,^e Helena Cortes Martins, MSc,^f Peter D. Kirwan, PhD,^{g,h} Kate O'Donnell, PhD,ⁱ Dimitrios Paraskevis, PhD,^a Cécile Sommen, PhD,^j Dominik Zenner, PhD,^k and Anastasia Pharris, PhD^d

Background: Migrant populations are overrepresented among persons diagnosed with HIV in the European Union and the European Economic Area. Understanding the timing of HIV acquisition (premigration or postmigration) is crucial for developing public health interventions and for producing reliable estimates of HIV incidence and the number of people living with undiagnosed HIV infection. We summarize a recently proposed method for determining the timing of HIV acquisition and apply it to both real and simulated data.

Methods: The considered method combines estimates from a mixed model, applied to data from a large seroconverters' cohort, with biomarker measurements and individual characteristics to

derive probabilities of premigration HIV acquisition within a Bayesian framework. The method is applied to a subset of data from the European Surveillance System (TESSy) and simulated data.

Findings: Simulation study results showed good performance with the probabilities of correctly classifying a premigration case or a postmigration case being 87.4% and 80.4%, respectively. Applying the method to TESSy data, we estimated the proportions of migrants who acquired HIV in the destination country were 31.9%, 37.1%, 45.3%, and 45.2% for those originating from Africa, Europe, Asia, and other regions, respectively.

Conclusions: Although the considered method was initially developed for cases with multiple biomarkers' measurements, its performance, when applied to data where only one CD4 count per individual is available, remains satisfactory. Application of the method to TESSy data, estimated that a substantial proportion of HIV acquisition among migrants occurs in destination countries, having important implications for public health policy and programs.

Key Words: HIV, migrants, surveillance systems, postmigration HIV acquisition

(*J Acquir Immune Defic Syndr* 2021;88:117–124)

Received for publication March 24, 2021; accepted June 2, 2021.

From the ^aDepartment of Hygiene, Epidemiology and Medical Statistics, National and Kapodistrian University of Athens, Medical School, Athens, Greece; ^bDepartment of Epidemiology of Infectious Diseases and Surveillance, National Institute of Public Health—National Institute of Hygiene, Warsaw, Poland; ^cStichting HIV Monitoring, Amsterdam, the Netherlands; ^dEuropean Centre for Disease Prevention and Control, Stockholm, Sweden; ^eInstitute for Global Health, University College London, London, United Kingdom; ^fNational Institute of Health, Lisbon, Portugal; ^gMedical Research Council Biostatistics Unit, University of Cambridge, Cambridge, United Kingdom; ^hBlood Safety, Hepatitis, Sexually Transmitted Infections and HIV Division, Public Health England, London, United Kingdom; ⁱHSE—Health Protection Surveillance Centre, Dublin, Ireland; ^jSanté Publique France, Saint-Maurice, France; and ^kCentre for Global Public Health, Institute for Population Health Sciences, Queen Mary University London, London, United Kingdom.

Supported by the European Centre for Disease Prevention and Control (Framework Contract ECDC/2018/014).

The authors have no conflicts of interest to disclose.

Conceived and designed the experiments: N.P. Analyzed the data: N.P. Wrote the article: N.P. and M.R. Drafted manuscript with critical revision for important intellectual content undertaken and contributed to interpretation of the data: A.V.S., C.Q., T.N., F.B., H.C.M., P.K., K.O.D., D.P., C.S., D.Z., and A.P. Verified the underlying data: N.P. and M.R. All authors gave final approval of the version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Correspondence to: Nikos Pantazis, PhD, Department of Hygiene, Epidemiology and Medical Statistics, National and Kapodistrian University of Athens, Medical School, 75 Mikras Asias, 115 27 Athens, Greece (e-mail: npantaz@med.uoa.gr).

Copyright © 2021 The Author(s). Published by Wolters Kluwer Health, Inc. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

INTRODUCTION

Migrant populations in the European Union (EU) and the European Economic Area (EEA) have been disproportionately affected by HIV. Although approximately 12% of people living in the EU were born in a different country from the one where they were a resident,¹ the corresponding percentage among those diagnosed with HIV in 2019 was 44%.²

A substantial proportion of HIV-positive migrants originate from countries with generalized epidemics. Almost 4 of 10 migrants in Europe diagnosed with HIV in 2019 were born in sub-Saharan Africa,² where HIV prevalence is substantially higher. This fact has contributed to the widespread belief that migrants are infected in their countries of origin, before arrival in Europe, although this viewpoint has been challenged.³

As the evidence accumulates, it becomes more and more clear that migrants are at an increased risk of many diseases postmigration, including sexually transmitted infections.^{4–7} Although migrants are typically a young and healthy population,⁸ limited access to HIV and sexually transmitted infections' prevention and testing, social inequalities, and structural factors

may lead to individual behaviors, which increase their vulnerability and may explain the higher HIV prevalence in comparison with native populations.^{4,5,9}

Knowing whether HIV acquisition takes place before or after migration has important public health policy implications because it impacts the design and delivery of HIV prevention and testing services.¹⁰ In addition, having estimates of the proportion of migrants who acquired HIV before entering the country where they were diagnosed is essential for model-based methods that estimate HIV incidence and the number of people living with HIV¹¹ and thus for the prospect of achieving epidemic control through UNAIDS 90-90-90 targets.

Unfortunately, discriminating between premigration and postmigration acquisition is difficult without specific HIV test results, such as a positive test before migration or conversely a negative test after migration or a positive recency test after migration. When only a single, positive HIV test is available, we can attempt to relate the probable time of infection to the time of migration. Systematic analysis to elucidate whether HIV acquisition took place in the country of origin or in the destination country has only been used by few European countries.^{3,12,13}

The methods that have been used to estimate the time of infection, at the individual level, are mainly based on biomarkers. If the average evolution of a biomarker during untreated HIV infection is well characterized, levels of that biomarker at or after an individual's diagnosis can be used to "back calculate" the elapsed time since infection. This principal idea has been applied to specific settings, often in combination with other auxiliary information (eg, life events, behavioral data, and other clinical information).^{3,12-20} These methods often require extended information not routinely available in surveillance data.

In this article, we adapt one of the most recent and flexible methods to estimate the time of infection¹⁹ and assess its performance using simulated data designed to resemble data sets collected routinely by European countries. Furthermore, we apply it to HIV surveillance data from the European Surveillance System (TESSy) of the European Centre for Disease Prevention and Control (ECDC) to generate estimates of the proportion of migrants infected premigration and postmigration in the EU/EEA.

METHODS

Data

We used data from TESSy,²¹ pooled in 2018, which included case reports of HIV diagnoses occurring between 1979 and the end of 2017 from 30 countries in the EU/EEA. Data from TESSy were provided by countries listed in the Acknowledgments and released by ECDC. We restricted the data to cases where age and sex were recorded; route of HIV transmission was through sex between men, heterosexual contact, or injecting drug use; and region of origin was not the same as the reporting country or unknown (N = 145,118). We excluded cases without any CD4 cell count measurement (32.4%) or where the year of arrival in the reporting country

(used as a proxy for migration date) was not available (33.8%). Finally, we removed cases where the date of HIV diagnosis was reported to be before the date of migration (4.2%), assuming that these could be classified as premigration infections. From the resulting subset of 43,005 cases, we randomly sampled 10% to illustrate the method in a midsized sample; thus, the working data for the current application included 4301 individuals.

For the simulation study, we generated a set of 5000 cases with known HIV seroconversion dates using the same methods as in the study of Pantazis et al.¹⁹ Simulated data approximated the subset of TESSy data used in the main study about certain characteristics: date of birth, sex, route of HIV transmission, region of origin, date of diagnosis, and migration date. Given that the focus of the current work is on the performance of the methods when applied to data from surveillance systems, we removed from the initially simulated data all HIV-RNA measurements, retained the first available CD4 cell count measurement, and excluded cases where AIDS occurred before the first available CD4 measurement or the route of transmission was other than the 3 categories considered in the TESSy data application.

Statistical Analysis

As for most biomarker-based methods, we first model the HIV natural history using data from a cohort of individuals with a known date of HIV seroconversion and applying a mixed model approach. Estimates from both fixed effects and variance components are then taken into account to formally derive not only just a point estimate of an individual's time of infection but also the distribution of this time.

More specifically, the method assumes that a bivariate linear mixed model, with known parameters, correctly characterizes the evolution of CD4 cell count and HIV-RNA viral load (appropriately transformed and denoted by superscripts *c* and *r*, respectively) over time since HIV seroconversion (*t*) while individuals are ART naive and AIDS free. Given the observed measurements, $y_i^T = (y_i^c, y_i^r)$, at times $t_i^T = (t_i^c, t_i^r)$ for the *i*-th individual, the distribution function can be written as $f(y_i^c, y_i^r | t_i^c, t_i^r)$. Denoting the observed times of markers' measurements relative to diagnosis by *d*, it follows that the time of the *j*-th measurement for the *i*-th individual since seroconversion can be expressed as $t_{ij}^c = d_{ij}^c + w_i$ and $t_{ij}^r = d_{ij}^r + w_i$ with the unknown quantity w_i denoting the time gap between HIV seroconversion and diagnosis date for the *i*-th individual. Thus, the distribution of the biomarkers, conditional on w_i , can be derived by replacing t_{ij}^c and t_{ij}^r with $d_{ij}^c + w_i$ and $d_{ij}^r + w_i$, respectively. Given the observed measurements, $y_i^T = (y_i^c, y_i^r)$, the posterior distribution of the unknown w_i conditional on y_i can be derived through the Bayes theorem as $f(w_i | y_i) = \frac{f(y_i | w_i) f(w_i)}{\int_0^{u_i} f(y_i | w_i) f(w_i) dw_i}$, $0 < w_i < u_i$ where u_i is the upper limit for the possible values of the gap between HIV seroconversion and diagnosis w_i (assuming that HIV seroconversion must have occurred after the age of 10, after 1/1/1980, and after any documented HIV negative test). A uniform prior distribution

for w_i , over the interval $(0, u_i)$, is assumed, but this distribution can be updated based on additional information.

The model $f(y_i^c, y_i^v | t_i^c, t_i^v)$ used to characterize the natural history evolution of CD4 cell count and HIV-RNA viral load was a bivariate linear mixed model. The model was fitted to the CASCADE seroconverters²² natural history data, in which seroconversion dates are well estimated. The following variables were used as covariates in the model: sex, age at HIV seroconversion, region of origin (Africa, Europe, Asia, or others), route of HIV transmission, and calendar year of HIV seroconversion. Category “others” in the region of origin variable mainly comprised Latin America and the Caribbean. Continuous covariates (year of seroconversion and age at seroconversion) entered the model through splines. CD4 decline was assumed linear after a square root transformation, and \log_{10} -transformed viral load was modeled through a fractional polynomial of time.

Given that the presence or absence of an AIDS-defining illness at or after HIV diagnosis carries additional information regarding the time gap between HIV infection and diagnosis, the formula for the posterior distribution of w_i can be accordingly updated. For example, a person diagnosed without AIDS would be more likely to have acquired HIV recently compared with a similar person diagnosed while having already progressed to AIDS. Thus, for an AIDS-free subject and not on therapy until some time d_i , since HIV diagnosis, the posterior distribution of w_i becomes $f(w_i | y_i, T_i > d_i + w_i) = \frac{f(y_i | w_i) S(d_i + w_i | X_i^s) f(w_i)}{\int_0^{m_i} f(y_i | w_i) S(d_i + w_i | X_i^s) f(w_i) dw_i}$ where T_i is a latent variable representing the time from HIV diagnosis to AIDS onset and $S(t | X_i^s)$ denotes the corresponding survival function conditional on subject-specific covariates.

This AIDS-free survival function $S(t | X_i^s)$ was estimated also using CASCADE data,²² truncated after 1/1/1996 (after that point, effective antiretroviral therapy which substantially reduced the probability of developing clinical AIDS became widely available) using a Weibull proportional hazards model with age at seroconversion and sex included as covariates.

Subject-specific estimates of the unknown gap between HIV seroconversion and diagnosis w_i , can be derived through the posterior mean, median, or mode, whereas the posterior probabilities of HIV acquisition postmigration for an individual, i , can be expressed as $\pi_i = P(w_i < m_i) = \int_0^{m_i} f(w_i | y_i) dw_i$ where m_i denotes the (known) time gap between migration and HIV diagnosis.

The method can be applied to individuals with multiple CD4 cell count and viral load measurements (provided they are taken before ART initiation or AIDS onset). However, it is also applicable when there is only one such measurement (CD4 or viral load) or even using information on AIDS status at diagnosis in case pre-ART and pre-AIDS CD4 cell count and viral load measurements are not available.

The estimation procedure was built in R (version 4.0.2)²³ using the base functions `optim` and `integrate` along with the additional library `mvtnorm` (version 1.1). For more details on the method, see reference 19, and for a previous application, see reference 3.

Role of the Funding Source

This work was supported by the European Centre for Disease Prevention and Control (Framework Contract ECDC/2018/014). Other than the coauthors from the ECDC who participated as individual experts, the study sponsor did not influence the study design, analysis, data interpretation, writing of the report, or the decision to submit the article for publication. The ECDC provided data from TESSy. A part of these data were used in analyses presented in this work.

RESULTS

Application to the Subset of TESSy Data

The subset of the TESSy data used for the current application included 4301 individuals. Their characteristics are summarized in Table 1. Most of the migrants (70.9%) in the sample originated from Africa (97% of whom were from sub-Saharan African countries), whereas smaller percentages originated from Europe (14.5%), Asia (5.4%), or other world regions (9.2%). European migrants in the sample originated equally from western (50.3%) and other European countries (49.7%).

Heterosexual contact was the predominant route of transmission (78.5%), followed by sex between men (19.5%) and injecting drug use (2.0%). The median time between arrival to the reporting country and diagnosis ranged from 2.3 to 4.7 years, for migrants originating from Europe and Asia, respectively. The median age at diagnosis was 34 years with European migrants being slightly younger (median 32 years). The median CD4 count at diagnosis was around 270 cells/ μ L for African or Asian migrants and around 410 cells/ μ L for migrants from Europe or other countries.

Most of the migrants (95.1%) in the sample had 1 CD4 cell count measurement available, taken soon after HIV diagnosis (93% within 3 months from diagnosis). A small proportion (4.4%) was diagnosed with clinical AIDS close to diagnosis and did not have any pre-ART/pre-AIDS CD4 cell count or HIV-RNA measurements. Finally, a very small proportion (0.5%) had more than 1 pre-ART/pre-AIDS CD4 cell count measurement or 1 CD4 cell count and 1 HIV-RNA viral measurement.

The results from the application of the method are summarized in Table 2. Estimated probabilities of premigration HIV acquisition were highest (0.64) for African migrants and 0.54 to 0.60 for migrants originating from other regions. Using the probability of 0.5 as a threshold, the percentages of migrants who were classified as having acquired HIV in the country of origin were 68.1%, 62.9%, 54.7%, and 54.8%, for those originating from Africa, Europe, Asia, and other regions, respectively. The median values for the estimated time between HIV acquisition and diagnosis ranged between 4.97 and 6.23 years, for migrants from other regions and Africa, respectively. Differences between the region of origin in all aforementioned estimates were highly statistically significant ($P < 0.001$).

Simulation Study

The simulated data set included 5000 cases contributing data on first CD4 count measurement after diagnosis and

TABLE 2. Results From the Application of the Method to the TESSy Subset Data

Region of Origin (Grouped)	Postmigration HIV Infection N (%)	Premigration HIV Infection N (%)	Probability of Premigration HIV Infection Mean (SD)	Time Between HIV Infection and Diagnosis, yr Median (IQR)
Africa	973 (31.9)	2076 (68.1)	0.64 (0.33)	6.23 (5.45–7.97)
Europe	232 (37.1)	394 (62.9)	0.60 (0.37)	5.30 (4.86–6.55)
Asia	105 (45.3)	127 (54.7)	0.54 (0.38)	5.81 (5.08–7.81)
Others	178 (45.2)	216 (54.8)	0.54 (0.38)	4.97 (4.50–6.06)
Overall	1488 (34.6)	2813 (65.4)	0.62 (0.35)	5.98 (5.19–7.63)

postmigration case) of 80.4%, with the positive and negative predictive values (ie, probabilities of being a premigration or postmigration case, given the corresponding classification by the method, respectively) being 91.0% and 73.7%, respectively.

Using the estimated probability of premigration infection as a continuous measure for discriminating between premigration and postmigration infections in an ROC analysis resulted in an area under the curve equal to 0.924 (Fig. 1).

Estimated time gaps between infection and diagnosis were close to the “true” ones with the difference in the overall median values being 0.24 years and no more than 0.5 years across different regions of origin. However, the agreement between estimated and observed time gaps from infection to diagnosis, for each case, was only moderate (the Lin coefficient of concordance = 0.409).

DISCUSSION

In this work, we adapted, evaluated in a simulation study, and applied to surveillance data a recently proposed

statistical method¹⁹ to estimate premigration or postmigration HIV acquisition among migrants to the EU/EEA. The simulation results showed a very satisfactory performance, whereas the application to data from European surveillance systems showed that a large proportion of HIV acquisitions occur postmigration.

This method was previously applied to a rich data set which included demographic and clinical variables, repeat biomarkers (CD4 cell counts and HIV-RNA measurements), and behavioral data within the aMASE study.³ Our current application of the method uses a subset of routinely collected data from TESSy²¹ in which the relevant information is much more restricted (eg, usually only one CD4 cell count, no HIV-RNA measurements, and no behavioral data). The accompanying simulation study was also tailored to mimic the data availability in typical surveillance systems.

The results from the simulation study showed that restricting the relevant data to variables collected in TESSy had an impact on the accuracy of the estimated time gaps between HIV acquisition and diagnosis. The concordance

TABLE 3. Characteristics of the Simulated Cases

	Region of Origin					P
	Africa	Asia	Europe	Others	Overall	
	N (%)	N (%)	N (%)	N (%)	N (%)	
Total	3660 (100.0)	296 (100.0)	652 (100.0)	392 (100.0)	5000 (100.0)	
Sex						<0.001
Male	1442 (39.4)	193 (65.2)	516 (79.1)	305 (77.8)	2456 (49.1)	
Female	2218 (60.6)	103 (34.8)	136 (20.9)	87 (22.2)	2544 (50.9)	
Route of transmission						<0.001
Heterosexual contact	3477 (95.0)	169 (57.1)	214 (32.8)	152 (38.8)	4012 (80.2)	
Sex between men	166 (4.5)	122 (41.2)	384 (58.9)	237 (60.5)	909 (18.2)	
Injecting drug use	17 (0.5)	5 (1.7)	54 (8.3)	3 (0.8)	79 (1.6)	
	Region of Origin					
	Africa	Asia	Europe	Others	Overall	
	Median (IQR)	Median (IQR)	Median (IQR)	Median (IQR)	Median (IQR)	P
Age at diagnosis, yr	33 (28–40)	32 (27–40)	31 (26–37)	32 (26–38)	33 (27–40)	<0.001
CD4 count at diagnosis, cells/μL	354 (197–608)	277 (127–498)	393 (234–646)	365 (203–592)	358 (198–604)	<0.001
Year of diagnosis	2008 (2004–2012)	2011 (2007–2014)	2010 (2006–2014)	2010 (2007–2014)	2008 (2004–2012)	0.005
Year of migration	2003 (1998–2008)	2005 (2000–2010)	2006 (2001–2011)	2005 (2001–2010)	2004 (1999–2009)	<0.001
Time from migration to diagnosis, yr	2.9 (0.8–7.1)	3.9 (1.5–8.2)	2.3 (0.7–5.6)	2.8 (0.9–6.5)	2.8 (0.9–6.9)	<0.001

TABLE 4. Results From the Application of the Method to the Simulated Data

Region of Origin (Grouped)		Postmigration HIV Infection N (%)	Premigration HIV Infection N (%)	Probability of Premigration HIV Infection Mean (SD)	Time Between HIV Infection and Diagnosis, yr Median (IQR)
Africa	Estimate	1196 (32.7)	2464 (67.3)	0.65 (0.32)	6.06 (5.33–7.43)
	True	1091 (29.8)	2569 (70.2)		6.46 (3.64–10.00)
Europe	Estimate	204 (31.3)	448 (68.7)	0.66 (0.32)	5.45 (4.89–6.64)
	True	194 (29.8)	458 (70.2)		5.12 (2.85–8.27)
Asia	Estimate	125 (42.2)	171 (57.8)	0.59 (0.34)	5.98 (5.13–7.94)
	True	112 (37.8)	184 (62.2)		6.46 (3.56–9.91)
Others	Estimate	142 (36.2)	250 (63.8)	0.61 (0.33)	5.23 (4.59–6.36)
	True	131 (33.4)	261 (66.6)		5.15 (3.09–8.37)
Overall	Estimate	1667 (33.3)	3333 (66.7)	0.64 (0.32)	5.95 (5.17–7.30)
	True	1528 (30.6)	3472 (69.4)		6.19 (3.48–9.64)
Correctly classified	Sensitivity	Specificity	PPV	NPV	Lin coefficient of concordance
	85.22%	87.36%	80.37%	91.00%	73.67%
					0.409

coefficient for simulated data mimicking the aMASE study¹⁹ was 0.69 and dropped to 0.40 in the current study. However, the percentage of cases which were correctly classified, based on the timing of HIV acquisition relative to date of migration, was 85.2% in the current simulation study, comparable with the aMASE study (84%).³ Similar findings hold for sensitivity and specificity of the method. These results are encouraging because they indicate that the method has good classification performance even in the absence of multiple biomarker measurements and behavioral data.

Applying the method to a subset of TESSy data, we estimated in a substantial proportion of migrants, HIV was acquired postmigration. The corresponding percentages were 31.9%, 37.1%, 45.3%, and 45.2% for those originating from Africa, Europe, Asia, and other regions, respectively. These estimates were lower than those found in the aMASE study (45% for African migrants and 69%–71% for migrants from other regions³). The difference between the 2 sets of results can be probably explained by differences in the selection of

the 2 samples: The aMASE study sample had to fulfill specific criteria (eg, diagnosed in the preceding 5 years, living in the destination country for at least 6 months, and only 9 European countries contributed data), whereas the subset of TESSy data includes diagnoses from a much wider time period over which the likelihood of premigration/postmigration HIV acquisition may have altered substantially. HIV cohorts, although broadly representative of the diagnosed population, were shown to underrepresent specific patient groups,²⁴ which may also affect the estimated proportion of premigration HIV acquisition. On the other hand, the high proportion of missing data on CD4 cell count and time of migration in TESSy resulted in a possibly selected subset which reflects the current data availability because these variables were generally more complete for the recent years of data. Thus, they may not be a representative of the whole EU/EEA migrant population diagnosed with HIV.

The main strength of the examined method is that it has a solid theoretical justification based on the full structure of a

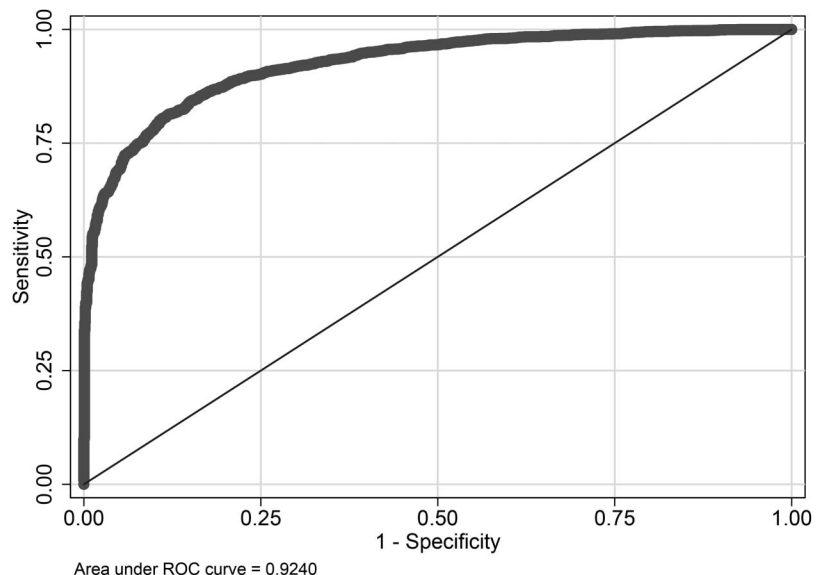


FIGURE 1. Receiver operating characteristic curve for the performance of the estimated probability of premigration infection.

mixed model for biomarker(s) evolution during the natural history of HIV. The method takes into account the evolution of the average biomarker levels over time and the evolution of their variance and covariance. In addition, the statistical and computational framework of the method allows the proper treatment of covariates which change over time (eg, age and calendar time), whose values are known at the time of diagnosis but unknown at the time of infection. The incorporation of additional external information is also feasible through informative prior distributions. Finally, the data used to fit the initial mixed model are derived from one of the largest collaborative seroconverter studies,²² which comprised 19,788 individuals contributing 125,195 CD4 cell count measurements, 106,160 HIV-RNA VL measurements, and a wide range of information on demographic and clinical characteristics.¹⁹

The main limitation of the examined method is that it assumes that, given a set of crucial covariates, the individuals we want to classify are similar to the seroconverters whose data were used to fit the initial mixed model. This is an assumption used in all biomarker-based methods, but a previous study²⁵ showed that estimates of HIV progression derived from seroconverters are likely to hold more generally for the HIV-positive population. Another drawback of the examined method is that it is more complex and computationally intensive compared with other methods which are based only on the evolution of the mean CD4 cell count and disregard the variance-covariance structure of the seroconverters' mixed model.^{13,15} In this regard, a tool to support the application of the proposed method by public health researchers working in HIV epidemiology would be useful.

Finally, it should be mentioned that our method does not take into account details of the migration process (eg, duration and route) or the probability of acquiring HIV during a short visit to the country of origin after migration. Behavioral data could be also used³ to improve the estimates of our method, but unfortunately, such data or data on the migration process are not routinely collected within surveillance systems. In general, the quality of data as collected in surveillance systems always remains a challenge, and in our model, especially, the poor completeness of CD4 count and time of arrival may lead to biased conclusions.

Besides the biomarker-based methods, another family of methods exists to assess premigration or postmigration HIV acquisition, based on molecular data analysis. For example, Paraskevis et al²⁶ used phylogenetic analysis to directly classify HIV infections as local or imported based on the idea of local transmission networks. Puller et al²⁷ presented a method to estimate the time gap between HIV infection and diagnosis based on the idea that diversity in the pol gene increases with time since infection. Finally, the concept of the molecular clock²⁸ could be applied as a method to elucidate the place of infection by estimating the time of infection and its relation to the time of migration. Unfortunately, molecular analysis methods require data and procedures that at this time are not routinely collected/performed within typical surveillance systems, although such data would be very important when studying and monitoring the HIV epidemic, and could be considered in surveillance in

the future. It should be noted that the examined biomarker method could incorporate information from molecular methods because of its Bayesian nature.

To conclude, the ability to discriminate between premigration and postmigration HIV acquisition is important for both monitoring the epidemic and designing testing and prevention strategies. The biomarker method we applied showed a very satisfactory ability to correctly classify diagnoses according to the place of HIV acquisition based on the information currently collected in the European surveillance system. The method is flexible and able to incorporate additional information (eg, from HIV-RNA viral load measurements or molecular data); thus, improvements in the quality and quantity of routinely collected data could further improve its performance. Most importantly, it is essential for surveillance systems to maintain and improve the collection of data on CD4 cell counts at diagnosis and migrants' time of arrival. Finally, our study also confirms that postmigration HIV acquisition is likely quite high and it merits routine monitoring in the future.

ACKNOWLEDGMENTS

The authors thank the members of the EU/EEA HIV Surveillance Network for their efforts to coordinate HIV surveillance at the national level and submission of data to the European Surveillance System. List of EU-EEA countries that provided data to TESSy are as follows: Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Sweden, and United Kingdom.

REFERENCES

1. Eurostat, population by age group, sex and country of birth. Available at: https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=migr_pop3ctb&lang=en. Accessed May 25, 2021.
2. ECDC. *HIV/AIDS Surveillance in Europe 2020*. Available at: <https://www.ecdc.europa.eu/sites/default/files/documents/hiv-surveillance-report-2020.pdf>. Accessed May 25, 2021.
3. Alvarez-Del Arco D, Fakoya I, Thomadakis C, et al. High levels of postmigration HIV acquisition within nine European countries. *AIDS*. 2017;31:1979–1988.
4. UNAIDS. *Gap Report*. Geneva, Switzerland: UNAIDS; 2014. Available at: <http://www.unaids.org/en/resources/campaigns/2014/2014gapreport/gapreport/>. Accessed May 25, 2021.
5. Alvarez-del Arco D, Monge S, Azcoaga A, et al. HIV testing and counselling for migrant populations living in high-income countries: a systematic review. *Eur J Public Health*. 2013;23:1039–1045.
6. Du H, Li X. Acculturation and HIV-related sexual behaviours among international migrants: a systematic review and meta-analysis. *Health Psychol Rev*. 2015;9:103–122.
7. Lewis NM, Wilson K. HIV risk behaviours among immigrant and ethnic minority gay and bisexual men in North America and Europe: a systematic review. *Soc Sci Med*. 2017;179:115–128.
8. Khat M, Darmon N. Is there a Mediterranean migrants mortality paradox in Europe? *Int J Epidemiol*. 2003;32:1115–1118.
9. Alvarez-Del Arco D, Monge S, Caro-Murillo AM, et al. HIV testing policies for migrants and ethnic minorities in EU/EFTA Member States. *Eur J Public Health*. 2014;24:139–144.
10. Fakoya I, Alvarez-del Arco D, Woode-Owusu M, et al. A systematic review of post-migration acquisition of HIV among migrants from countries with generalised HIV epidemics living in Europe: implications

- for effectively managing HIV prevention programmes and policy. *BMC Public Health*. 2015;15:561.
11. van Sighem A, Nakagawa F, De Angelis D, et al. Estimating HIV incidence, time to diagnosis, and the undiagnosed HIV epidemic using routine surveillance data. *Epidemiology*. 2015;26:653–660.
 12. Brannstrom J, Sonnerborg A, Svedhem V, et al. A high rate of HIV-1 acquisition post immigration among migrants in Sweden determined by a CD4 T-cell decline trajectory model. *HIV Med*. 2017;18:677–684.
 13. Rice BD, Elford J, Yin Z, et al. A new method to assign country of HIV infection among heterosexuals born abroad and diagnosed with HIV. *AIDS*. 2012;26:1961–1966.
 14. Berman SM. A stochastic model for the distribution of HIV latency time based on T4 counts. *Biometrika*. 1990;77:733–741.
 15. Desgrees-du-Lou A, Pannetier J, Ravalihasy A, et al. Sub-Saharan African migrants living with HIV acquired after migration, France, ANRS PARCOURS study, 2012 to 2013. *Euro Surveill*. 2015;20:46.
 16. Drylewicz J, Commenges D, Thiébaud R. Maximum a posteriori estimation in dynamical models of primary HIV infection. *Stat Commun Infect Dis*. 2012;4:10.
 17. Geskus RB. On the inclusion of prevalent cases in HIV/AIDS natural history studies through a marker-based estimate of time since seroconversion. *Stat Med*. 2000;19:1753–1769.
 18. Munoz A, Carey V, Taylor JM, et al. Estimation of time since exposure for a prevalent cohort. *Stat Med*. 1992;11:939–952.
 19. Pantazis N, Thomadakis C, Del Amo J, et al. Determining the likely place of HIV acquisition for migrants in Europe combining subject-specific information and biomarkers data. *Stat Methods Med Res*. 2019;28:1979–1997.
 20. Gosselin A, Ravalihasy A, Pannetier J, et al. When and why? Timing of post-migration HIV acquisition among sub-Saharan migrants in France. *Sex Transm Infect*. 2020;96:227–231.
 21. *The European Surveillance System (TESSy)*. 2021. Available at: <https://www.ecdc.europa.eu/en/publications-data/european-surveillance-system-tesy>. Accessed May 25, 2021.
 22. *CASCADE (Concerted Action on Seroconversion to AIDS and Death in Europe)*. Available at: <https://www.ctu.mrc.ac.uk/studies/all-studies/c/cascade/>. Accessed May 25, 2021.
 23. *R: A Language and Environment for Statistical Computing. R Core Team*. 2019. Available at: <https://www.R-project.org/>. Accessed May 25, 2021.
 24. Vourli G, Pharris A, Cazein F, et al. Are European HIV cohort data within EuroCoord representative of the diagnosed HIV population? *AIDS*. 2019;33:133–143.
 25. Lodi S, Phillips A, Touloumi G, et al. CD4 decline in seroconverter and seroprevalent individuals in the precombination of antiretroviral therapy era. *AIDS*. 2010;24:2697–2704.
 26. Paraskevis D, Kostaki E, Nikolopoulos GK, et al. Molecular tracing of the geographical origin of human immunodeficiency virus type 1 infection and patterns of epidemic spread among migrants who inject drugs in Athens. *Clin Infect Dis*. 2017;65:2078–2084.
 27. Puller V, Neher R, Albert J. Estimating time of HIV-1 infection from next-generation sequence diversity. *Plos Comput Biol*. 2017;13:e1005775.
 28. Leitner T, Albert J. The molecular clock of HIV-1 unveiled through analysis of a known transmission history. *Proc Natl Acad Sci U S A*. 1999;96:10752–10757.