Are European HIV cohort data within EuroCoord representative of the diagnosed HIV population?

Georgia Vourlia, Anastasia Pharris, Francoise Cazein, Dominique Costagliola, Francois Dabis, Julia Del Amo, Valerie Delpech, Asuncion Díaz, Enrico Girardi, Annabelle Gourlay, Barbara Gunsenheimer-Bartmeyer, Victoria Hernando, Georgios Nikolopoulos, Kholoud Porter, Magdalena Rosińska, Caroline Sabin, Barbara Suligojn, Virginie Supervie, Ferdinand Wit and Giota Touloumi

Objective: HIV cohorts are an important source of clinical data for informing public health policies and programmes. However, the generalizability of cohort findings to the wider population of people diagnosed with HIV in each country remains unclear. In this work, we assessed the representativeness of six large national HIV cohorts within Europe.

Design and methods: Individual-level cohort data were provided from national cohorts in France, Germany, Greece, Italy, Spain and the United Kingdom. Analysis focused on new HIV diagnoses reported to The European Surveillance System (TESSy) during three time periods (2000–2004, 2005–2009 and 2010–2013), to allow for temporal changes. Cohort and TESSy records were matched and compared by age, sex, transmission mode, region of origin and CD4⁺ cell count at diagnosis. The probability of being included in each cohort given demographic characteristics was estimated and used to generate weights inversely proportional to the probability of being included.

Results: Participating cohorts were generally representative of the national HIV-diagnosed population submitted to TESSy. However, people who inject drugs, those born in

---

*Correspondence to Dr Georgia Vourli, PhD, Department of Hygiene, Epidemiology and Medical Statistics, Medical School, NKUA, Athens, Greece. 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

HIV infection continues to be of major public health importance worldwide, notably because significant numbers of new infections and premature deaths continue to occur [1]. In the era of test and treat approaches and of biomedical prevention [2], the development and assessment of effective public health responses requires the monitoring of not only new diagnoses and routes of transmission, but also other key markers including linkage to care, treatment uptake and viral suppression.

HIV surveillance in Europe is carried out by the European Centre for Disease Prevention and Control (ECDC) and the WHO Regional Office for Europe. HIV cases reported annually by European Union Member States to The European Surveillance System (TESSy), hosted at ECDC, remain the best estimate of the number of people diagnosed with HIV, although underreporting does occur in some Member States. Although the demographic data collected by TESSy provides important information for monitoring, data on clinical indicators are often lacking. EuroCoord was a Network of Excellence established by several of the largest HIV cohorts and collaborations within Europe – as part of this collaborative effort, EuroCoord supported the development of a combined dataset including individual demographic data and clinical indicators for almost 300,000 HIV-positive adults and children [3].

HIV cohorts are an important source of data for informing policy. However, cohorts may systematically exclude specific groups of individuals, those less likely to access care, such as people who inject drugs (PWID) [4–6]. The representativeness of each cohort is usually assessed by comparing the characteristics of cohort participants to those of corresponding reference populations [7–10]. Several methods to account for selection bias have been proposed [11–13].

The aim of the present work was to assess the representativeness of several large cohorts of HIV-positive adult individuals followed within EuroCoord by comparing their characteristics with those diagnosed from the same countries as submitted to TESSy. These findings can inform our understanding of whether, and in which settings and population groups, results based on cohort data can be generalized to the wider population, allowing ECDC, European Union Member States, and the European Commission to confidently use cohort data to provide guidance on issues of public health importance with regard to HIV infection.

Methods

Seven cohorts within EuroCoord that do not form the basis of their respective national surveillance systems were included in analyses: the French Hospitals Database on HIV ANRS CO4 and Aquitaine ANRS CO3 from France [14,15], Clinical Surveillance of HIV Disease (ClinSurv) from Germany [16], Athens Multicentre AIDS Cohort Study (AMACS) from Greece, Italian Cohort Naïve Antiretrovirals (ICONA) from Italy [17], Cohort of the Spanish HIV Research Network (CoRIS) from Spain, and the UK Collaborative HIV Cohort (UK CHIC) from the United Kingdom [18].

TESSy data of individuals diagnosed up to 2013 and reported up to 2015 in these six countries were provided by ECDC under agreement for data confidentiality. Anonymized data with variables on age, sex, transmission mode, country/region of origin, year of diagnosis and CD4⁺ cell counts at diagnosis were included. Individual cohort data on the same variables were provided from each participating country. All participating cohorts had approval from relevant country authorities.

Statistical analysis

TESSy data were considered as the gold standard against which the characteristics of cohort participants were compared. Analyses focused on HIV cases diagnosed during the time periods 2000–2004, 2005–2009 and 2010–2013 to account for changes in individual characteristics over time. Each country and time period was analysed separately.
To avoid confusion due to different percentages of missing data in TESSy and the cohorts, multivariate imputation by chained equations was applied before analysis (see Appendix, http://links.lww.com/QAD/B370) [19]. Transmission mode was categorized as: MSM, PWID and ‘other’, which mainly included those reportedly infected through heterosexual contact. Region of origin was categorized as: reporting country, other European (nonreporting) country/Australia/North America and Africa/Asia/Latin America. Age was categorized as 15–24.9, 25–34.9, 35–44.9, 45–54.9 and at least 55 years; and CD4⁺ cell counts as less than 200, 200–349, 350–499 and at least 500 cells/µL. Region of origin was not adequately recorded in ClinSurv, whereas ethnicity instead of origin was recorded in the UK CHIC Study and thus it was not analysed in these countries. In ICONA, participants from Italy were recorded as Europeans, and thus the region of origin was not adequately recorded in ClinSurv, whereas ethnicity instead of origin was recorded in the UK CHIC Study and thus it was not analysed in these countries. In ICONA, participants from Italy were recorded as Europeans, and thus the region of origin could only be categorized in two groups. CD4⁺ cell counts were not reported to TESSy from Germany and for France were only reported from 2008 onwards.

TESSy and cohort datasets from each country were merged by a pseudo-id created by combining each individual covariate pattern with a unique within-covariate pattern serial number. The distributions of individual characteristics in the two datasets were compared. An indicator of whether a diagnosed individual reported to TESSy was included in the cohort was generated. Multivariable logistic regression models were fitted to estimate the probability of diagnosed individuals being included in the cohort given their characteristics and stabilized weights inversely proportional to this probability were generated [11–13].

Inference was based on the median stabilized weights for the levels of the studied covariates, pooled over all periods and other covariates, along with the corresponding interquartile range (IQR); underrepresentation required that the entire IQR was greater than 1.

Results

In total, 235 437 individuals reported to TESSy and 95 768 cohort participants of known sex, diagnosed after 2000 and during the periods covered by both surveillance and cohorts, were included in the analyses.

In general, the cohorts captured the main characteristics of the HIV-diagnosed individuals in the corresponding countries (Table 1). Main trends over time observed in TESSy were also seen in the cohorts. However, specific subgroups were underrepresented (Fig. 1). PWID, those born in another country, those with low CD4⁺ cell counts at diagnosis, older individuals and women were, in general, underrepresented in the cohorts, but in different degrees across different countries (Fig. 1). PWID were underrepresented in France, Greece, Spain and the United Kingdom. In Germany, this finding held only for the 2005–2009 period. In Italy, Spain and the United Kingdom, individuals reporting other transmission mode were underrepresented; in France and Greece individuals originated from other than the reporting country and in Spain those originated from another (nonreporting) European country were also underrepresented; individuals who were older than 55 years were underrepresented in Italy, Spain, the United Kingdom and, to some extent, in Greece. Women were also slightly underrepresented in most countries including Italy, Spain and the United Kingdom, whereas in Greece, women were underrepresented only in the 2005–2009 period.

Although the degree of underrepresentation or overrepresentation was quite low when we considered each variable separately, it varied substantially when we considered the degree of representation by covariate patterns (i.e. combining all variables). In France, PWID originated from outside the country were highly underrepresented (stabilized weights: 1.8–13.1), as were MSM originated from Africa, Asia or Latin America (stabilized weights: 1.2–4.6). Similarly, in Greece, PWID originated from other European countries were substantially underrepresented (stabilized weights: 1.6–4.9). In Spain, PWID originated from other countries were also underrepresented (stabilized weights: 1.5–3.4).

Application of inclusion weights efficiently compensated for observed imbalances and reproduced the structure of the corresponding diagnosed population (Appendix, Fig. A1, http://links.lww.com/QAD/B370).

Discussion

This is the first effort to assess the representativeness of European HIV cohorts within EuroCoord compared with the wider population of HIV-diagnosed individuals reported to TESSy. Our findings suggest that, whilst cohort participants are a fairly representative sample of HIV-diagnosed individuals, some groups remain underrepresented or overrepresented within these cohorts. All participating cohorts, other than CoRIS in which participants are also recruited from an STI clinic, are HIV clinic-based. Thus, differences between cohorts and TESSy data may reflect differences between individuals who are and are not linked to care. PWID tended to be underrepresented in all cohorts and time periods, except in ICONA. Impaired access to care and retention in care for PWID is a global challenge [6,20,21]. Older individuals also tended to be underrepresented in some cohorts. Results from previous studies are conflicting, reporting either higher or lower probability of inclusion.
Table 1. Demographic characteristics in cohort and The European Surveillance System data for individuals diagnosed between 2000 and 2013, Germany, Greece and the United Kingdom; for France, only individuals diagnosed between 2003 and 2013 were included and for Italy only data for 2010–2013, whereas in Spain only individuals diagnosed between 2005 and 2013 are included.

<table>
<thead>
<tr>
<th>Transmission mode</th>
<th>France</th>
<th>Germany</th>
<th>Greece</th>
<th>Italy</th>
<th>Spain</th>
<th>United Kingdom</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cohort, N = 28 758</td>
<td>Surveillance, N = 59 480</td>
<td>Cohort, N = 14 146</td>
<td>Surveillance, N = 33 845</td>
<td>Cohort, N = 33 033</td>
<td>Surveillance, N = 15 449</td>
</tr>
<tr>
<td>MSM</td>
<td>11 067 (38.5)</td>
<td>20 238 (34.0)</td>
<td>8206 (58.0)</td>
<td>4535 (53.5)</td>
<td>6232 (40.5)</td>
<td>5239 (62.6)</td>
</tr>
<tr>
<td>PWID</td>
<td>577 (2.0)</td>
<td>2305 (3.9)</td>
<td>948 (6.7)</td>
<td>1482 (17.5)</td>
<td>947 (6.1)</td>
<td>497 (5.9)</td>
</tr>
<tr>
<td>Other</td>
<td>17 114 (59.5)</td>
<td>36 937 (62.1)</td>
<td>4992 (35.3)</td>
<td>2458 (29.0)</td>
<td>8250 (53.4)</td>
<td>2638 (31.5)</td>
</tr>
<tr>
<td>Sex</td>
<td>Male</td>
<td>19 091 (66.4)</td>
<td>38 144 (64.1)</td>
<td>11 191 (79.1)</td>
<td>27 691 (81.8)</td>
<td>11 898 (77.0)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>9667 (33.6)</td>
<td>21 336 (35.9)</td>
<td>2955 (20.9)</td>
<td>6154 (18.2)</td>
<td>6988 (42.0)</td>
</tr>
<tr>
<td>Age group</td>
<td>15–24.9</td>
<td>3316 (11.5)</td>
<td>5543 (9.3)</td>
<td>1671 (11.8)</td>
<td>3953 (11.7)</td>
<td>731 (11.1)</td>
</tr>
<tr>
<td></td>
<td>25–34.9</td>
<td>10 109 (35.2)</td>
<td>18 632 (31.3)</td>
<td>4476 (31.6)</td>
<td>11 614 (34.3)</td>
<td>788 (9.3)</td>
</tr>
<tr>
<td></td>
<td>35–44.9</td>
<td>8639 (30.0)</td>
<td>18 762 (31.5)</td>
<td>4777 (33.8)</td>
<td>10 681 (31.6)</td>
<td>1487 (17.5)</td>
</tr>
<tr>
<td></td>
<td>45–54.9</td>
<td>4346 (15.1)</td>
<td>10 613 (17.8)</td>
<td>2098 (14.8)</td>
<td>5074 (15.0)</td>
<td>1487 (17.5)</td>
</tr>
<tr>
<td></td>
<td>&gt;55</td>
<td>2348 (8.2)</td>
<td>5930 (10.0)</td>
<td>1124 (7.9)</td>
<td>2523 (7.5)</td>
<td>1487 (17.5)</td>
</tr>
<tr>
<td>Origin</td>
<td>Reporting country</td>
<td>18 889 (65.7)</td>
<td>29 138 (49.0)</td>
<td>9575 (67.7)</td>
<td>18 039 (53.3)</td>
<td>5228 (84.1)</td>
</tr>
<tr>
<td></td>
<td>Europe, USA</td>
<td>731 (2.5)</td>
<td>2576 (4.3)</td>
<td>1595 (11.3)</td>
<td>2619 (7.7)</td>
<td>413 (6.6)</td>
</tr>
<tr>
<td></td>
<td>Africa, Asia, etc</td>
<td>9138 (31.8)</td>
<td>27 766 (46.7)</td>
<td>2976 (21.0)</td>
<td>13 187 (39.0)</td>
<td>576 (9.3)</td>
</tr>
<tr>
<td>CD4+ cells/µL</td>
<td>&lt;200</td>
<td>1309 (23.5)</td>
<td>6186 (30.3)</td>
<td>1702 (27.4)</td>
<td>2948 (34.8)</td>
<td>881 (26.7)</td>
</tr>
<tr>
<td></td>
<td>200–349</td>
<td>1291 (23.1)</td>
<td>4433 (21.7)</td>
<td>1351 (21.8)</td>
<td>1901 (22.4)</td>
<td>616 (19.3)</td>
</tr>
<tr>
<td></td>
<td>350–499</td>
<td>1196 (21.5)</td>
<td>3804 (18.7)</td>
<td>1191 (19.2)</td>
<td>1399 (16.5)</td>
<td>666 (20.2)</td>
</tr>
<tr>
<td></td>
<td>&gt;500</td>
<td>1785 (32.0)</td>
<td>5968 (29.3)</td>
<td>1971 (31.7)</td>
<td>2227 (26.3)</td>
<td>1120 (33.9)</td>
</tr>
</tbody>
</table>

PWID, people who inject drug.

*aFor France CD4+ cell counts concern only on the last study period (2008–2013).*
for older individuals [5,22]. However, it has been reported previously that older age is significantly associated with late diagnosis [23]. As previously described [5], women were also slightly underrepresented in most cohorts. In accordance with previous studies that report suboptimal access to care among migrants, there were indications that migrants were underrepresented [24,25]. Individuals with low CD4\(^+\) cells were less likely to be included in most cohorts, possibly because they may die shortly after their diagnosis, or because they often start combined antiretroviral treatment (cART) immediately and are perhaps excluded from cohorts recruiting cART-naïve individuals (e.g. ICONA, CoRIS). Underrepresentation of individuals with a low CD4\(^+\) cell count in cohorts has also been reported [5].

In this work, TESSy data were considered the gold standard with which cohort data were compared. However, surveillance data also have limitations. Changes over time were taken into account by analysing each study period separately. To account for reporting delays, we analysed data up to 2013, whereas the average delays in the analysed countries are less than 6 months [26]. Surveillance data included in TESSy may underestimate the number of people diagnosed with HIV in some countries, because of underreporting (as in France, underreporting was estimated at 29.6% for 2010–2013 in France). Additional limitations include a lack of data on migrant status as well as misclassified and/or missing data that are more likely to occur in surveillance settings than in cohorts. Thus, differences between cohort and surveillance data may be due to the different conditions under which data collection is performed by cohorts and surveillance systems.

The observed differences may also reflect delayed enrolment into the cohorts. In cases of recent epidemics in specific groups, enhanced linkage to care is essential so that these newly infected individuals are promptly provided access to ART and clinical follow-up and, where relevant, included in HIV cohorts.

In summary, our findings indicate that the main European cohorts capture a representative sample of the HIV-
diagnosed population. Nevertheless, the sample of individuals participating in HIV cohorts within EuroCoord may differ systematically from the population of HIV-diagnosed individuals reported to ECDC through TESSy. Given that vulnerable patients tend to be underrepresented, estimation of public health indices based on cohort data could lead to overoptimistic conclusions. Provided that cohorts are able to capture data on all characteristics associated with inclusion into the cohort, weighted analyses are likely to provide more unbiased estimates of public health-indices [12,13]. Supplementing surveillance data with annual updates of public health relevance such as treatment uptake and viral suppression should be the gold standard of HIV surveillance. This information can be provided from existing clinical cohorts and should ideally be expanded to all patients accessing HIV care in a given country. Therefore, the results of this project could be used to more effectively triangulate HIV surveillance and cohort data for public health action.

Acknowledgements

FHDH-ANRS CO4 cohort, France.

The FHDH ANRS CO4 cohort is funded by the ANRS, INSERM and the French Ministry of Health.


Corevih Bretagne (CHU de Rennes: C. Arvieux, F. Souala, J.M. Chapplin, M. Ratajczak, J. Rohan), Corevih Centre Poitou-Charentes (CHRU de Tours), Corevih Franche-Comté (CH de Belfort: J.P. Faller, O. Ruyer, V. Gendrin, L. Toko; CHRU de Besançon: C. Chirouze, L. Hustache-Mathieu, J.F. Faucher, A. Proust, N. Magy-Bertrand, H. Gil, N. Méaux-Ruault); Corevih Haute-Normandie (CHRU de Rouen), Corevih Languedoc-Roussillon (CHU de Montpellier; CHU de Nîmes; A. Sotto, I. Rouanet, J.M. Maubourquin, R. Doncesco, G. Jacques), Corevih Lorraine Champagne Ardennes (Nancy-Hôpital de Brabois: T. May, C. Rabaud, M. Andre, M. Delestan, M.P. Bouillon; CHRU de Reims: F. Bani-Sadr, C. Rouger, J.L. Berger, Y. Nguyen), Corevih de Midi-Pyrénées Limousin (Toulouse CHU Purpan: B. Marchou, P. Delobel, G. Martin Blondel, L. Cuzin, N. Biezunski, L. Alric, D. Bonnet, M. Guivarch, A. Palacin, V. Paysan), Corevih Nord-Pas de Calais (CH de Tourcoing: H. Melliez, F. Ajana, A. Meybeck, N. Vignet), Corevih PACA Est (Nice Hôpital Archet 1: C. Pradier, P. Pugliese, P.M. Roger, E. Rosenthal, J. Durant, E. Cua, A. Naqvi, I. Perbost, K. Risso; C.H. Antibo: D. Quinsat; CHI de Lille: C. Rouger, J.L. Berger, Y. Nguyen), Corevih PACA Ouest (Marseille-Hôpital de la Conception: P. Ramb, S. Antolimi-Bouvenot, P. Druart, L. Meddeb, I. Ravaux, A. Menard, C. Tomei, C. Dhiyer, H. Tissot-Dupant; Marseille-Hôpital Nord: J. Moreau, S. Mokhtari, M.J. Soavi, V. Thomas; Marseille-Hôpital Sainte-Marguerite: I. Poizot-Martin, S. Breggeon, O. Faucher, V. Obry-Roguet, A.S. Ritleng, N. Petit; Marseille-Centre pénitentiaire des Baumettes: C. Bartoli, J.M. Ruiz, D. Blanc; C.H. d’Aix-En-Provence: T. Allegre, M. Sordage, J.M. Riou, C. Faudon; C.H. d’Avignon: B. Slama, H. Zerazhi, O. Viallard J.F.); Infectious Diseases and Tropical Medicine, Pellegrin Bordeaux University Hospital, (Cazanave C., Dauchy C., Dupon M., Dutronc H., Neau D., Malvy D., Pistone T., Receveur M.C., Wirth G.); Internal Medicine and Infectious Diseases, Haut Leveque Bordeaux University Hospital, (Greib C., Lazaro E., Pellegrin JL., Viallard J.F.); Internal Medicine, Agen Hospital, (Imbert Y., Rispal P.); Internal Medicine, Libourne Hospital, (Caubet O., Ferrand H., Tchamgoue S.); Infectious Diseases, Bayonne Hospital, (Farbos S., Vareil M.O., Wille H.); Internal Medicine and Infectious Diseases, Dax Hospital, (Andre K., Caunegre L., Gerard Y.); Infectious Diseases, Saint-Cyr Hospital, (Chossat L.); Internal Medicine, Mont de Marsan Hospital (Gerard Y.); Internal Medicine, Arcachon Hospital, (Coulpault C.); Internal Medicine and Infectious Diseases, Perigueux Hospital, (Berthol N., Cougoul B., Latape P., Marie J., Rouanes N.); Internal Medicine and Infectious Diseases, Pau Hospital, (Dumondin G., Gaborieau V.); Internal Medicine, Orthez Hospital, (Gerard Y.).

ClinSurv HIV, Germany.

The clinical surveillance of HIV, ClinSurv HIV, is funded by the Robert Koch Institute, which is the German Public Health Institute.
The current work was supported by the Spanish Network of HIV/AIDS (RD12/0017/0018 and RD16CIII/0002/0006) and CIBER Epidemiología y Salud Pública (CIBERESP), Spain. J. del Amo, V. Hernando, A. Díaz: ‘This article presents independent results and research. The views expressed are those of the authors and not necessarily those of the Instituto de Salud Carlos III’.


UK CHIC, United Kingdom.

UK CHIC is funded by the UK Medical Research Council (Grant numbers G0000199, G0600337, G0900274 and M004236). The views expressed in this article are those of the researchers and not necessarily those of the Medical Research Council.


Authors’ contribution: G.V.: performed the statistical analysis and drafted the article. A.P., A.G., K.P., M.R. and F.W. acted as advisors for the study design and the statistical analysis. D.C., F.D., J.D.A., V.D., E.G., B.G.B., V.H., C.S., V.S., G.V. and G.T. described in detail eligibility criteria, geographical coverage and all characteristics of the cohorts they represented for this work. J.D.A., V.D., C.S., D.C. were also member of the study advisory committee. F.C., A.D., B.G.B., V.H. and B.S.
discussed the corresponding characteristics of their countries’ National surveillance systems. G.T. supervised the study design and analysis and coauthored the draft article. All authors contributed substantially in finalizing the article.

Conflicts of interest
A.P. is employed by the ECDC. D.C. reports grants from Janssen-Cilag (2017–2018), Merck-Sharp & Dohme-Chibret (2015–2017), ViiV (2015), personal fees from Janssen-Cilag (2016, 2018) and Merck-Sharp & Dohme-Chibret (2015, 2017) for lectures, personal fees from ViiV (2015), for travel/accommodations/meeting expenses, personal fees from Gilead Science, Janssen, Otsuka Novel Products and Angelini for consultancy or lectures outside the submitted work. G.T. has received grants unrelated to this study from Abbvie (2018) and Janssen (2018), outside the submitted work. G.T. has received grants unrelated to this study from Abbvie (2018) and Janssen (2018), outside the submitted work. G.T. has received grants unrelated to this study from Abbvie (2018) and Janssen (2018), outside the submitted work. G.T. has received grants unrelated to this study from Abbvie (2018) and Janssen (2018), outside the submitted work. G.T. has received grants unrelated to this study from Abbvie (2018) and Janssen (2018), outside the submitted work. G.T. has received grants unrelated to this study from Abbvie (2018) and Janssen (2018), outside the submitted work.

References