

G. Vourli¹, A. Pharris², F. Cazein³, D. Costagliola⁴, F. Dabis⁵, J. Del Amo⁶, V. Delpech⁷, A. Díaz⁶, E. Girardi⁸, A. Gourlay⁹, B. Gunsenheimer-Bartmeyer¹⁰, V. Hernando⁶, G. Nikolopoulos^{11,12}, K. Porter⁹, M. Rosińska¹³, C. Sabin⁹, B. Suligoi¹⁴, V. Supervie⁴, F. Wit¹⁵, G. Touloumi¹

¹ Department of Hygiene, Epidemiology and Medical Statistics, Medical School, NKUA, Athens, Greece. ² European Centre for Disease Prevention and Control (ECDC), Stockholm, Sweden. ³ Institut national de Santé publique France, Saint-Maurice, France. ⁴ Sorbonne Universités, UPMC, INSERM, IPLESP UMR_S 1136, France. ⁵ Université Bordeaux, ISPED, Centre INSERM U1219 - Bordeaux Population Health, Bordeaux, France. ⁶ National Center of Epidemiology, Instituto de Salud Carlos III, Madrid, Spain. ⁷ Public Health England, London, UK. ⁸ Clinical Epidemiology Unit, National Institute for Infectious Diseases L. Spallanzani, Rome, Italy. ⁹ Department of Infection and Population Health, Faculty of Population Health Sciences, University College London, London, UK. ¹⁰ Robert Koch-Institute Dep. of Infectious Disease Epidemiology HIV/AIDS, STI Unit, Berlin, Germany. ¹¹ Hellenic Centre for Disease Control and Prevention, Amaraouso, Greece. ¹² Medical School, University of Cyprus, Nicosia, Cyprus. ¹³ Department of Epidemiology, National Institute of Public Health-National Institute of Hygiene, Warsaw, Poland. ¹⁴ Centro Operativo AIDS, Istituto Superiore di Sanità, Rome, Italy. ¹⁵ Stichting HIV Monitoring, Amsterdam, the Netherlands.

Introduction

- ECDC collects demographic data such as age group, gender, transmission mode and country of origin of HIV diagnosed individuals through TESSy, but data on clinical indicators are often lacking.
- Within EuroCoord, the European Network of HIV/AIDS Cohort Studies to coordinate clinical research on HIV/AIDS, data on clinical indicators are collected (e.g drug resistance, viral suppression, level of compliance to cART initiation guidelines, survival)
 - This data could be used to supplement data collected through TESSy
- Challenges include issues associated with cohort and surveillance data:
 - Cohorts typically include HIV diagnosed individuals who were linked to care, a selected sample that may be different from the whole HIV diagnosed population
 - Cohorts do not cover the whole country geographically or possibly systematically exclude specific group(s) of patients (e.g. IDU, migrants), or are otherwise restricted (e.g. requiring patient's consent)
 - Surveillance systems may have substantial changes over time regarding their geographical coverage, suffer from underreporting, missing data and usually lack of data on outcome (death) or outmigration of cases notified historically
- Within EuroCoord:
 - 5 countries have a cohort-based national surveillance system, thus no issue of representativeness.
 - 6 countries include data from a subset of HIV diagnosed individuals. Analysis focused on these countries.

Aims

- Assess the representativeness of data on HIV patients of cohorts within EuroCoord against persons diagnosed with HIV in corresponding countries and reported to TESSy.
- Improve understanding of whether and where cohort data can be generalized
- Explore and propose methods to improve cohorts' representativeness

- Cohorts' coverage in the participating countries varied from 21% to 83% between 2000 and 2013 (Figure 1). The decline in the recent years indicates either delayed cohorts' data update or improvement in corresponding HIV surveillance systems.
- According to logistic regression models people injecting drugs, those born in another country and those with low CD4 counts at diagnosis were less likely to be included in almost all cohorts. Women and older individuals were also under-represented occasionally.
- The distribution of the stabilized weights according to region of origin, transmission mode and CD4 category is illustrated in Figure 2. Stabilised weights above 1 suggest under-representation in the cohort.
- Application of inclusion weights in an example cohort is illustrated in Figure 3. Although there are differences in the distribution of individual characteristics in cohort and surveillance data, variables' distributions approaches the corresponding distributions in TESSy data after applying the inclusion weights.

Results

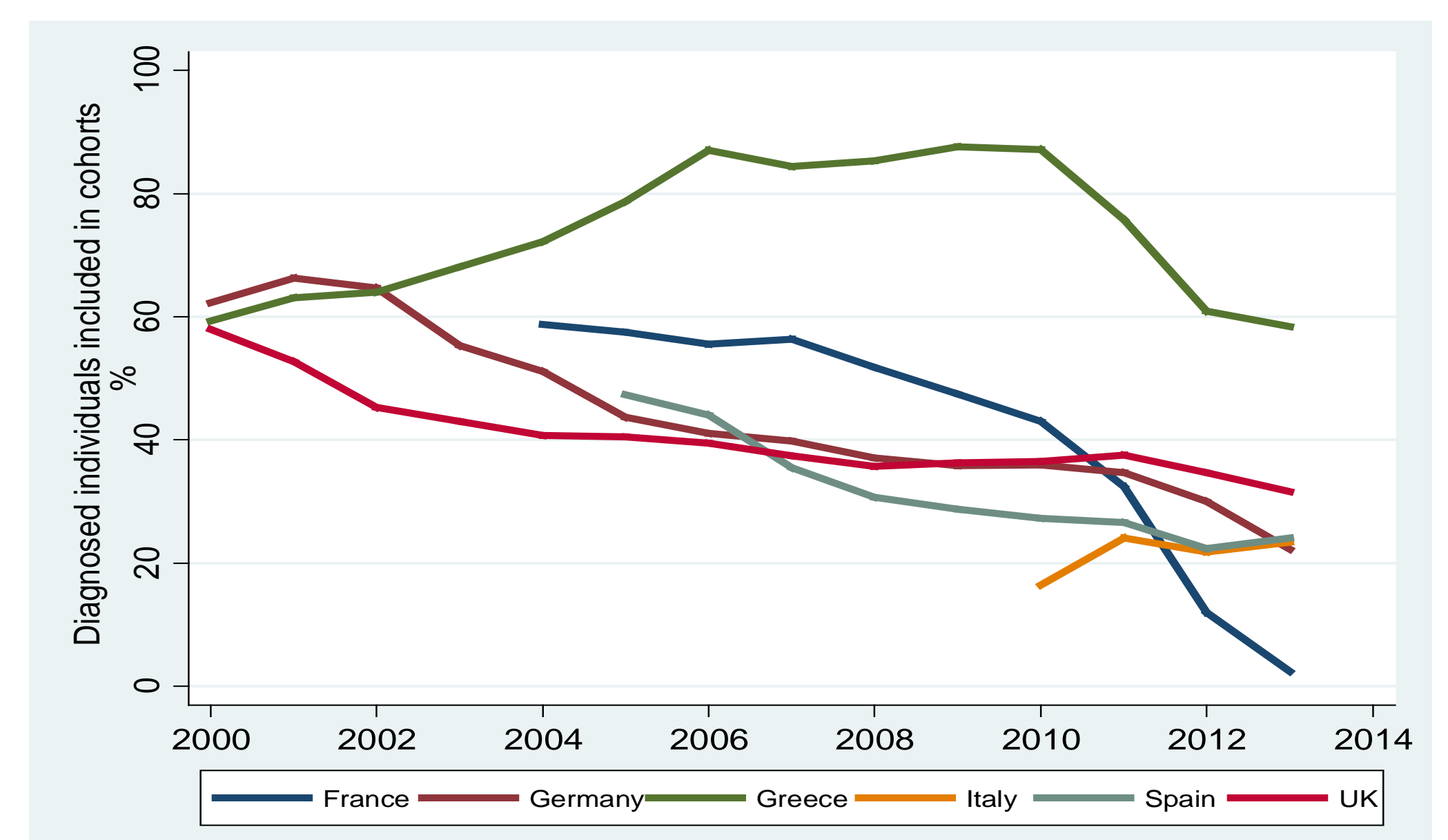


Figure 1. Percentage of diagnosed individuals included in the cohorts, from 2000 to 2013.

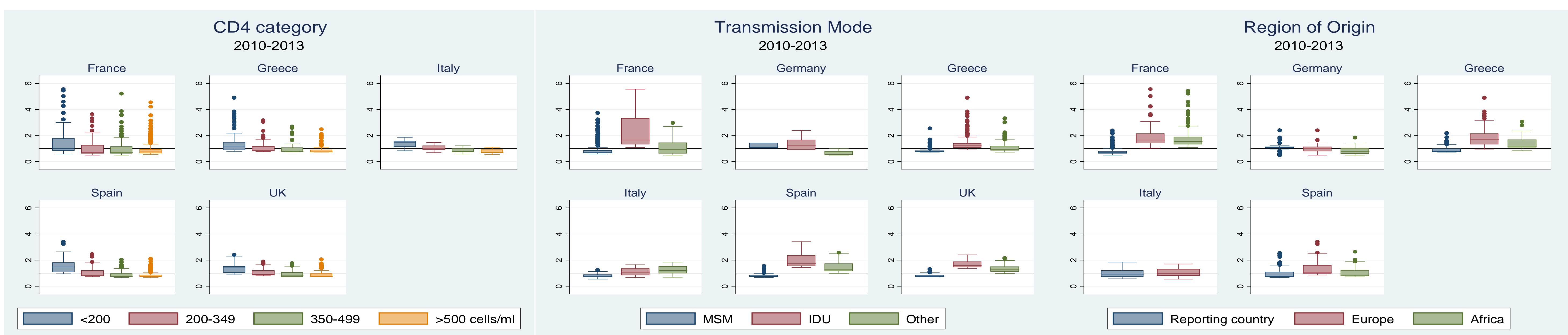


Figure 2. Stabilised weights according to region of origin, transmission mode and CD4 category. Horizontal line at 1 corresponds to an ideal representation of individuals in the cohorts, while values above 1 correspond to under-represented groups of patients.

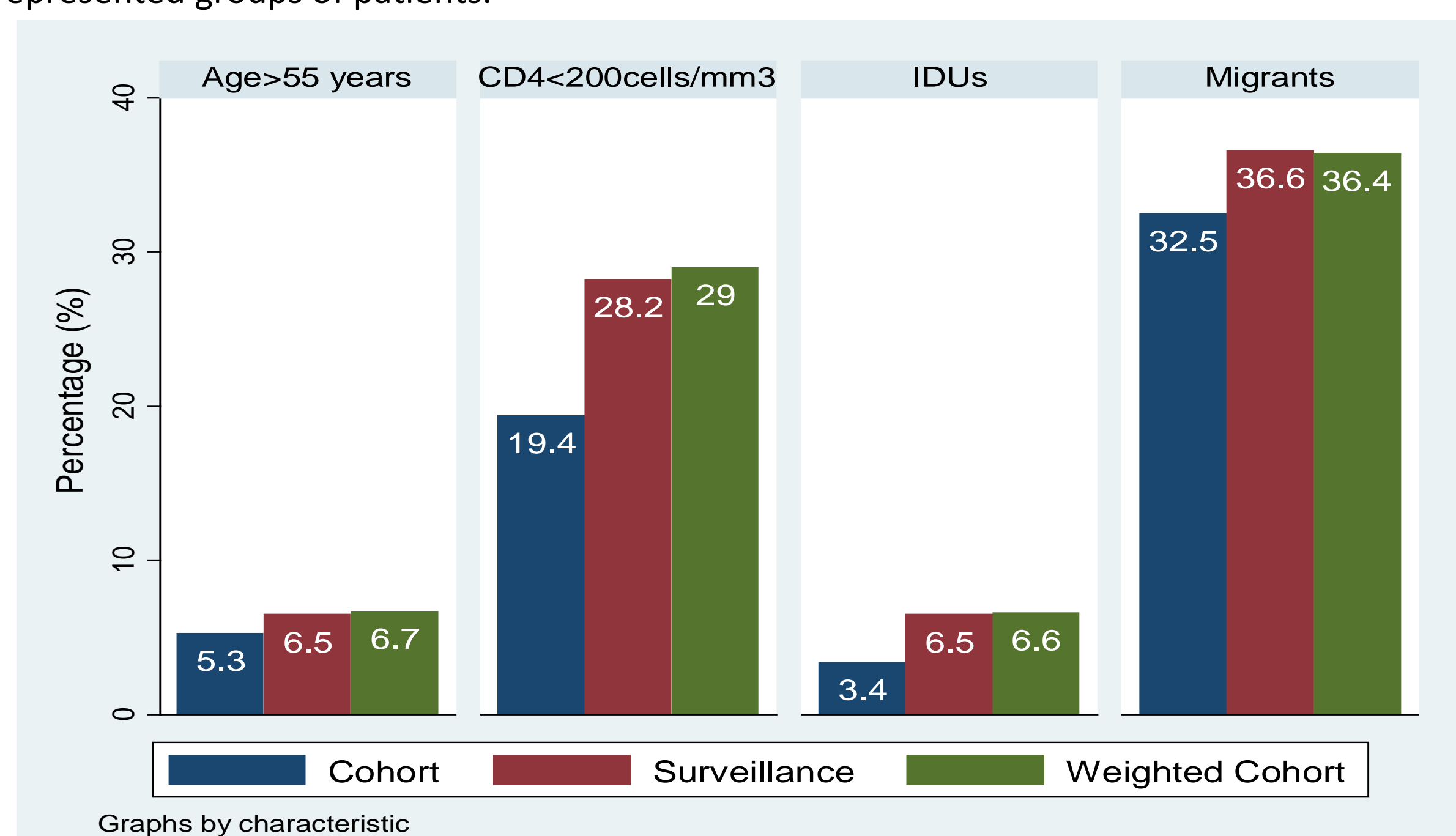


Figure 3. Distributions of basic characteristics in the cohort, surveillance data and after applying the inclusion weights (Spain, 2010-2013).

Methods

- France, Germany, Greece, Italy, Spain and the UK provided individual cohort data.
- To accommodate countries' specific features, new cases diagnosed during the three time periods [2000-2004], [2005-2009] and [2010-2013] were analysed separately.
- Distribution of individuals' age at diagnosis, gender, transmission mode and region of origin were compared between cohorts and corresponding TESSy data.
- Logistic regression models were fitted to estimate the probability of diagnosed individuals being included in the cohort given their demographic characteristics.
- Weights inversely proportional to the probability of inclusion were generated for each covariate pattern and assigned to each cohort participant as:

$$weight = \frac{1}{Pr(\text{inclusion in the cohort} | \text{individual characteristics})}$$

- Weights can be interpreted as the number of copies each cohort participant should contribute to reproduce the corresponding population of diagnosed individuals.
- Stabilised weights were also produced, to facilitate inference:

$$\text{stabilised weight} = \frac{Pr(\text{inclusion in the cohort})}{Pr(\text{inclusion in the cohort} | \text{individual characteristics})}$$

- The denominator represents the cohort's coverage. Thus, the stabilised weights are distributed around 1, with a value <1 indicating over-representation and a value >1 under-representation in the cohort
- To mimic the structure of the underlying population of HIV diagnosed individuals, under-represented subgroups are assigned greater weights, whereas over-represented subgroups are assigned smaller weights.

Conclusions

- Limitation: Issues concerning the surveillance systems (e.g. under-reporting, reporting delays) were not taken into account
- European cohorts within EuroCoord capture a fairly representative sample of the corresponding population of HIV diagnosed individuals
- Vulnerable HIV diagnosed individuals (IDUs, migrants, diagnosed with low CD4) are most likely to be under-represented in the cohorts
- Weighting can be applied to correct for mis-representation of subgroups of patients in the cohorts
- Results of this project could be used to more effectively triangulate HIV surveillance and EuroCoord data for public health action
- Weights can be applied to other analyses and cohorts