Working package 6

Analysis of the variation across Europe of educational differential in activity limitation

EU-SILC 2009

INED

EUROPEAN HEALTH & LIFE EXPECTANCY
JA:EHLEIS
INFORMATION SYSTEM

The EHLEIS team comprises:

Jean-Marie Robine, INSERM U988 and U710, Montpellier, France, jean-marie.robine@inserm.fr
Herman Van Oyen, Scientific Institute of Public Health, Brussels, Belgium, Herman.VanOyen@wiv-isp.be
Nicolas Berger, Scientific Institute of Public Health, Brussels, Belgium, Nicolas.Berger@wiv-isp.be
ŠárkaDaňková, Institute of Health Information and Statistics of the Czech Republic, Praha, Czech Republic, dankova@uzis.cz.
Bernard Jeune, University of Southern Denmark, Institute of Public Health, Odense, Denmark, BJeune@health.sdu.dk
Henrik Bronnum Hansen, Department of Public Health, University of Copenhagen, Denmark, Henrik.Bronnum-Hansen@sund.ku.dk
Ola Ekholm, National Institute of Public Health, Copenhagen, Denmark, oek@niph.dk
Barbara Hjalsted, Danish National Board of Health, Copenhagen, Denmark, bah@sst.dk
Mikkel Baadsgaard, Economic Council of the Labour Movement, Copenhagen, Denmark, mb@ae.dk
Emmanuelle Cambois, INED (Institut National d'Etudes Démographiques), Paris, France, Cambois@ined.fr
France Meslé, INED, (Institut National d'Etudes Démographiques), Paris, France,mesle@ined.fr
Isabelle Mougenot, University of Montpellier II, Montpellier, France, mougenot@lirmm.fr
Gabriele Doblhammer, Rostock Center for Demographic Change, Germany, doblhammer@rostockerzentrum.de
Jürgen Thelen, Robert Koch Institute, Berlin, Germany, ThelenJ@rki.de
Lars Kroll, Robert Koch Institute, Berlin, Germany, L.Kroll@rki.de
Giorgos Ntouros, Hellenic Statistical Authority, Athens, Greece, Geodouro@statistics.gr
Viviana Egidi, University la Sapienza, Rome, Italy, viviana.egidi@uniroma1.it
Wilma J. Nusselder, Erasmus Medical Center, Rotterdam, Netherlands, w.nusselder@erasmusmc.nl
Caspar Looman, Erasmus Medical Center, Rotterdam, Netherlands, c.looman@erasmusmc.nl
Hendriek Boshuizen, National Institute for Public Health and the Environment, Bilthoven, Netherlands,Hendriek.Boshuizen@rivm.nl
Jan-Willem Bruggink, Statistical Office (CBS), Heerlen, Netherlands, jw.bruggink@cbs.nl
Marten Lagergren, National Board of Health and Welfare (SoS/NBHW), Stockholm, Sweden, marten.lagergren@aldrecentrum.se
Carol Jagger, Newcastle University, United-Kingdom, carol.jagger@newcastle.ac.uk
Chris White, Office of National Statistics, Newport, United-Kingdom, Chris.White@ons.gsi.gov.uk
Tony Fouweather, Newcastle University, United-Kingdom, tony.fouweather@newcastle.ac.uk
Stefaan Demarest, Scientific Institute of Public Health, Brussels, Belgium, stefaan.demarest@wiv-isp.be
Denise Walckiers, Scientific Institute of Public Health, Brussels, Belgium, denise.walckiers@wiv-isp.be
KaatjeBollaerts, Scientific Institute of Public Health, Brussels, Belgium, kaatje.bollaerts@wiv-isp.be
Leila Oumeddour, INED (Institut National d'Etudes Démographiques), Paris, France, leila.oumeddour@ined.fr
Isabelle Beluche, INSERM U710/CRLC, Montpellier, France, isabelle.beluche@inserm.fr
Cedric Hassen-Khodja, CRLC, Montpellier, France, cedric.hassen-khodja@montpellier.unicancer.fr
Christine Perrier, CRLC, Montpellier, France, christine.perrier@montpellier.unicancer.fr

Contact EHLEIS:
Jean Marie ROBINE, INSERM
Centre Val d'Aurelle,
Parc Euromédecine,
34298 Montpellier cedex 5, France.
Tel: +33 (0) 467 61 30 29
Fax: +33 (0) 467 61 37 87
Email: jean-marie.robine@inserm.fr

Joint action EHLEIS co-funded by DG SANCO (Agreement number 20102301)

Contents

I. The EU-SILC survey, variables of interest and data robustness 5

a. The EU-SILC project 5

b. Variables: Activity Limitation and Education 5

c. Completeness of variables and representativity of national samples 6

II. Health inequalities in EU-SILC 10

a. Prevalences, inequalities and hypothetical groups 10

b. National prevalence of levels 5-6 as a function of relative risk 15

i. Countries' relative positions 15

ii. Clustering using the k-means method 15

c. Logistic regression 19

I. The EU-SILC survey, variables of interest and data robustness

a. The EU-SILC project

The aim of the EU-SILC (European Union – Statistics on Income and Living Conditions) is to collect comparable cross-sectional and longitudinal data from across Europe on the themes of income, poverty, social exclusion and living conditions. EU-SILC was launched in 2003 in 6 Member States. The 2009 version of the project concerns 576,208 individuals, distributed among 29 European countries (the then-27 Member States along with Norway and Iceland). This instrument is based on the European Statistical System (ESS): its guiding framework is provided by the commission (Eurostat) and it is applied in each country by the national statistical institutes (NSI). EU-SILC includes several questions on health, notably the Global Activity Limitation Indicator (GALI) which is used to calculate the European health expectancy indicator (Healthy Life Years, or HLY)

The aim of the work presented here is to contribute to studying the possibility of building indicators of the relationship between HLY and socioeconomic status, in order to estimate the magnitude of health disparities within the European Union. This report presents the initial results of an analysis of health inequalities by level of education, after the introduction of all variables of interest as well as information on data robustness.

b. Variables: Activity Limitation and Education

Our study aims to investigate health inequalities by level of education, which was made possible in SILC 2009 by the use of the variables "Activity Limitation" and "Education":

Activity Limitation - GALI

The Global Activity Limitation Indicator (GALI) refers to any state of health, illness, mental or physical problem, which has limited the respondent's ability to perform usual actions or tasks for at least the last 6 months (it concerns limited activity itself and not the health problems that may precede it). Three possible responses are proposed: 1) Yes, strongly limited; 2) Yes, limited; 3) No, not limited. For this variable, Eurostat recommends limiting the use of proxies as much as possible. This is the variable that is used to calculate HLY, and which we will focus on here.

Education – Highest level of education attained

This variable represents the highest level of education attained. Its coding corresponds to the seven highest levels of the UNESCO ISCED (International Standard Classification of Education) framework, in its 1997 version, and not the most recent version, adopted in 2011. Since its inception, the data in the EU-SILC survey have included only six categories, from 0 to 5, the last of these grouping together levels 5 and 6 from the UNESCO framework. A definition of each level can be found in the Appendix. [1]

[1] ISCED (International Standard Classification of Education) en Anglais

In an article discussing the reliability of the analysis of levels of education in Europe in the EU-SILC survey,[2] Silke Schneider and Walter Müller recommended aggregating these six categories into three in order to limit the effects of very large variations in the numbers of individuals in the different categories. This grouping limits the analytical scope of the education variable, but that scope had already been substantially reduced upstream: by providing information only on the most aggregated levels, the original coding makes it impossible to constitute subgroups based on other criteria (such as grouping together those with general vs. vocational training). According to Schneider and Müller, this aggregation produces categories which are relatively stable across the different waves of the survey, as well as in comparison with the Eurostat EU-LFS (Labour Force Survey) data that we use here. These groups are used for our analyses.

Grouping recommended by Schneider and Müller

Pre-primary education	0
Primary education	1
Lower secondary education	2
Upper secondary education	3
Post-secondary non-tertiary education	4
First stage of tertiary education	5
Second stage of tertiary education	6

c. Completeness of variables and representativity of national samples

Because the data was standardized after collection, contrary to many international surveys, the various countries in the survey used diverse methodologies, be it in terms of data collection or of sample adjustment. We have identified four data collection methods (Table 1). Countries had the option either of collecting information through interview-based surveys with a sample of the population or drawing on national statistical sources which fulfill Eurostat data quality criteria. Nineteen countries opted to collect almost all data through face-to-face interviews, either CAPI (computer-assisted personal interviews) or PAPI (paper-and-pencil interviews); three others drew on a mixture of face-to-face and telephone interviews. Nine countries (the Netherlands, Sweden, Norway, Denmark, Finland, Slovenia, Iceland, Ireland and Latvia) drew a part of the data on households and individuals from their national registers, complementing it with interviews for the collection of individual data: in Ireland and Latvia data was mainly collected through face-to-face interviews; in Slovenia, half of the data was collected face-to-face and half by telephone; and the majority was collected by telephone in the other six countries. Finally, in Denmark 7% of interviews were given as self-administered questionnaires, and in Germany all interviews were conducted in this way.

In certain countries, there were large losses between the initial SILC base and the final sample, including all individuals for whom health inequalities by level of education could be analyzed. However, this loss did not systematically alter the quality of observations. This depends on the

[2] Schneider S, Müller W. "Measurement of Education in EU-SILC – Preliminary Evaluation of Measurement Quality", Equasolc Working Paper 2009/5.
http://www.equalsoc.org/publication/283. (21 November 2011).

determinants of the loss, which, if they are linked to health and/or social status, can change the prevalence of activity limitations in the surveyed groups: reducing it if non-respondents are more limited (refusal/inability to participate) or increase it if, on the contrary, non-respondents are less limited (employed young people who are difficult to reach). Two effects are at work simultaneously: a selection effect, if non-respondents and non-participants differ significantly from respondents in terms of health or level of education; and the distortion effect that can modify the prevalence of the population as a whole, if the structure of the sample differs significantly from that of the population under study, shifting the mean value toward that of the most represented category.

The selection effect associated to non-participants in EU-SILC is not measurable, insofar as their social status and health are unknown. The selection effect associated to non-respondents in the final sample can be assessed insofar as partial information, notably on social status, can be filled in from the household database even if the individual did not participate in the individual survey. The distortion can be assessed by comparing the social structure of the final sample with that of the complete EU-SILC database and the reference data provided by Eurostat.

While for 13 countries the sample structure is close to that of the population as a whole, in other cases the two differ to varying degrees. In certain countries low levels of education are under-represented, in some cases quite markedly (SE, IS, MT, UK) and in others less so (CY, ES, FR, GR, NL, SK). In the Belgian sample there was also a tendency toward over-representation of the most educated category, but the proportion of respondents in the least educated category is the same as in the population. In Luxembourg, the least educated group is substantially over-represented; there as in Iceland and Malta, sample distortions lead to the under- or over-representation of one of the education categories by more than 10 percentage points. For a few countries, the least educated category is over-represented (EE, IE, LT).

In terms of the age structure of the sample, there is a tendency for the youngest age group to be under-represented; the determinants of non-response to the individual survey mainly being youth and employment (non-reachability), probably an effect of conducting many individual interviews by telephone. Note, however, that information on education, occupation and health is also quite frequently lacking for individuals above the age of 80. This result is partly explained by the fact that among the elderly, responses may be more often provided by a proxy who, on the one hand, may be uninformed about the person's level of education, and who on the other hand is not authorized in many countries to respond to health question on the person's behalf.

Overall, in countries that used registers and telephone surveys, participation rates in the individual survey and completeness levels of the education and health variables were very low. This high sample loss does not, however, seem to have significantly altered the structure of the sample with respect to age, level of education or socio-professional category. We note, then, an under-representation of the youngest group (unreached) and the oldest (when responses were given by a proxy): the result must be an underestimation of activity limitation in the 80+ group, insofar as there is a high probability at these ages that response by proxy is related to difficulties responding directly; in the youngest group, the effect of a response by proxy or of non-participation must more often be linked to lack of time or to refusal to participate, without any corresponding incapacities. The direction of the effect thus probably depends on age, but also on social status: the groups that are under-represented among the elderly probably lead to the under-estimation of the prevalence of

activity limitations, and the effect on the prevalence of the population will be all the greater for the fact that this under-representation disproportionately applies to the social groups where activity limitations are most prevalent, namely among those with the lowest level of education.

Given their structure, it is preferable not to include the countries which substantially under- or over-represent certain groups: this includes Malta, Luxembourg and Iceland. Moreover, to limit uncertainties, in this study we will focus on ages 30-79, where information is more complete. More generally, specifying the countries with suspected bias is recommended, considering the deformation of the samples' social structure.

Table 1: Participation rate, sample size and percentage of individuals recorded by each method of data collection, by country

	Individual participation rate	EU-SILC	Final sample	Loss	Data collection method	
					Registers	Interview method
AT	71.10%	13610	11054	19%		CAPI / CATI
BE	62.70%	14721	11651	21%		CAPI / CATI
BG	77.20%	15047	13148	13%		PAPI
CY	89.50%	9283	7553	19%		CAPI / CATI
CZ	82.30%	23302	16827	28%		PAPI / CAPI
DE	76.50%	28368	23686	17%		Self-admin
DK	53.50%	15025	5866	**61%**	x	CATI / Self-admin
EE	74.00%	13542	11220	17%		PAPI / CAPI
ES	81.00%	36865	30418	17%		CAPI / CATI
FI	79.20%	25157	9962	**60%**	x	CAPI / CATI
FR	82.70%	25611	20113	21%		CAPI
GR	84.00%	18035	15045	17%		PAPI / CAPI / CATI
HU	84.50%	25053	20354	19%		PAPI
IE	78.90%	12641	9900	22%	x	CAPI
IS	73.10%	8545	2895	**66%**	x	CATI
IT	83.70%	51196	42159	18%		PAPI
LT	86.90%	12852	10700	17%		PAPI / CAPI
LU	51.90%	11406	8491	26%		PAPI
LV	78.30%	14403	12066	16%	x	PAPI / CAPI / CATI
MT	79.80%	10213	8478	17%		CAPI
NL	83.40%	23687	9717	**59%**	x	CATI
NO	60.40%	13855	5349	**61%**	x	CAPI / CATI
PL	76.30%	38541	29228	24%		PAPI
PT	86.40%	13013	11091	15%		PAPI / CAPI
RO	96.20%	18703	16282	13%		PAPI
SE	73.00%	18441	7540	**59%**	x	CATI
SI	77.70%	29576	9276	**69%**	x	CAPI / CATI
SK	88.50%	16137	13636	15%		PAPI
UK	71.30%	19380	15359	21%		CAPI

Table 2: Problems encountered for each country and at each step in the analysis of the quality of the data

	Individual Response Rate	Final Sample Loss	Missing Data			Representativity			Distortions			Variations in prevalences		Use of substitutes
			Education	Occupation	Health	Education	Occupation	Age	Education	Occupation	Age	Limitation	Severe limitation	
AT	*													
BE	**			*			*							
BG	*													
CY							*							
CZ		*						*	*	*	*			
DE	*						*							
DK	**	**	80+ **	*					**	**	**			
EE	*				**									
ES						*								**
FI		**							**	**	**			
FR				*, 80+ *										
GR						*								
HU														
IE						*		*				**		**
IS	*	**				***			*	*	**			
IT														
LT					**									
LU	**					***	(**)					**		
LV	*													
MT	*			80+ *		***								
NL		**		*					*	*	**			
NO	**	**		**			*		*	*	*			
PL	*		*								*			
PT														
RO														
SE	*	**	80+ *	*		*			*	**	**			
SI	*	**							*	**	*			
SK							*	*						
UK	*		*			*								

II. Health inequalities in EU-SILC

Our aim is to study health inequalities in Europe on the basis of the data from the 2009 version of the SILC survey.

a. Prevalences, inequalities and hypothetical groups

The graphs (Figures 2 and 3) show the national prevalences of the participant countries in SILC (except Iceland, Malta and Luxembourg, where the analysis of data collection casts doubt on the reliability of the 2009 survey data). These bubble charts show the prevalences of the groups with the highest and lowest levels of education (5-6 and 0-2) for each country, total national prevalence, and the corresponding 95% confidence intervals. The size of each bubble corresponds to the relative size of the corresponding educational group. The prevalences of the most educated groups as well as relative risk are given for each country. The countries are classified into two large groups according to their date of entry into the European Union (entry before or upon the 5th enlargement of 2004-2007), as well as their national prevalence.

We thus present a graph for each expanded age category: 30-49, 50-64 et 65-79, standardized. We then present the same graphs, but this time using hypothetical groups. These were created by harmonizing education groups such that each group contains a third of the surveyed individuals. These new distributions assume that prevalence is a linear function of educational level. On the basis of the regression line, the prevalences of the three education groups, if each were to include a third of individuals, can be determined. These new hypothetical groups offer a way of limiting selection bias due to the differing size of the education groups. The groups themselves being aggregations of different levels of education, a small 0-2 group may reflect a homogeneous group concentrating a high number of individuals with a very low level of education, which could be associated to high prevalence. A larger 0-2 group could be linked to more heterogeneous education levels, corresponding in this case to lower prevalence.

These hypothetical groups can also be used to calculate a hypothetical risk ratio, corresponding to the increased risk of experiencing activity limitation for the least educated third of the population with respect to the most educated third of the population. Individuals were classified into terciles: the x-axis gives the cumulative percentages of individuals by level of education, and the y-axis the corresponding prevalences (*P*), calculated after standardizing the education groups by age.

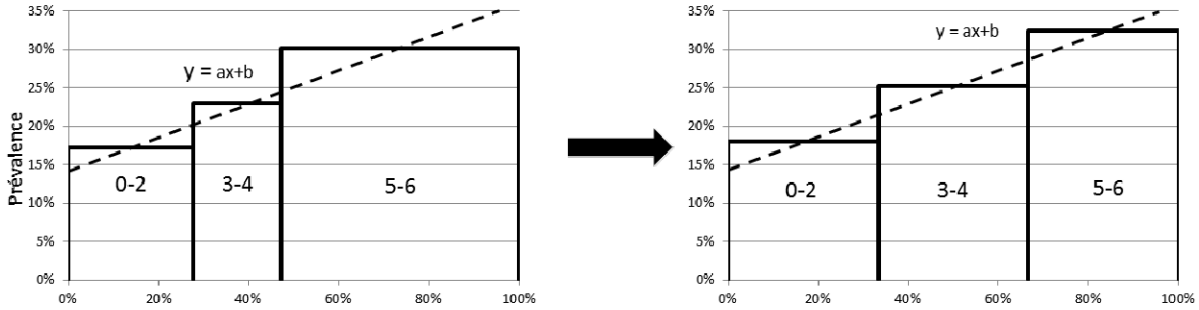
$$RR = P0 - 2P5 - 6$$

The hypothetical ratio calculated on the basis of these new prevalences thus takes into account the size of the different education groups in each country and normalizes them, allowing international comparisons. Relative risks and hypothetical relative risks are shown in Figures 2 and 3.

The gradient of national prevalence levels in the EU15 countries (with NO and excluding LU) is slightly more favourable than that of the EU12 countries (excluding MT), although they substantially overlap: the first runs from 35% (DE) to 16% (SE), while the second runs from 41% (SK) to 17% (BG). The prevalences of the most educated run from 26% (DE) to 10% (NO) and from 34% (SK) to 10%

(CY), and those of the least educated from 46% (DE) to 22% (SE) and from 54% (SK) to 20% (BG). Relative risks run from 1.4 (FI) to 2.9 (NO) for the EU15 and from 1.5 (BG) to 2.8 (CY).

Figure 1: Creation of hypothetical education groups on the basis of the regression line



These graphs also highlight the importance of the relative size of the different education groups in the calculation of national prevalence, which is "attracted" by large bubbles: in the countries where a high proportion of the population falls into the least educated group, national prevalence is closer to that of this group (PT, ES, IT, GR, IE / RO, CY), and the opposite is true in countries where a high proportion of the population is in the most educated group (DK, NO, SE / SK EE). For the other countries, high proportions in the intermediate education group attract prevalence, which thus falls in between the levels of the two extreme groups, whose bubbles are small.

The creation of hypothetical groups artificially compensates for these distributional effects and makes trans-national comparisons of the gap between classes possible. After this adjustment, relative risks run from 1.4 (FI) to 2.7 (NO) for the EU15 and from 1.4 (BG) to 2.6 (CY) for the EU12. The adjustment significantly modified relative risk in a number of countries: by a maximum of around 0.10 point in the countries of the EU15 (DE, AT, IT, GR, NO, SE), but particularly in the EU12 countries where the risk ratios decreased in many cases by more than 0.2 point (HU, SI, LT, RO), and even considerably more (in the case of CZ). Composition effects thus had a strong impact not only on mean prevalence but also on the magnitude of differences, according to whether the extreme groups represented a large population, or else a small and thus selected population.

For the youngest generations, national prevalence is generally strongly influenced by the highest levels of education (excepting PT, IT and ES), whereas for the older generations, national prevalence is more attracted by lower levels of education.

For the 30-49 cohorts, the national prevalences and prevalences of the most educated groups of the EU12 countries (excluding MT) are lower overall than in the other countries, generally associated to larger 0-2 education groups. However, there are large disparities, with national prevalences in each group of countries going from around 5% (GR) to approximately 20% (FI, DK, DE, NL, SK). Moreover, the 0-2 education group's size seems to be correlated to its prevalence: greater relative size seems to be linked to lower prevalence. With progressively older generations, the prevalences of the EU12 countries increase more than those of the other countries: At ages 65-79, whereas national prevalences run from around 25% (NO) to 55% (DE) for the countries that entered the EU earlier, those of the new entrants run from a little over 35% (BG) to nearly 80% (SK).

The prevalences of education groups with small sample sizes have larger dispersions, in particular 0-2 education groups at ages 30-49. The representation of hypothetical prevalences makes it possible to compensate for these large dispersions, which are particularly pronounced in the youngest cohorts for low levels of education, and in the oldest cohorts for high levels of education. This representation also makes possible trans-national comparisons of the prevalences by educational level.

For almost all countries, hypothetical relative risk ratios are lower than observed relative risks. Only Belgium, for ages 30-79 and more particularly 30-49, Spain, for ages 30-79, and Estonia, for ages 30-49, show a significant increase in relative risk after the construction of the hypothetical groups. Note that as a general rule, the decreases seem to be larger for the group of new entrants than for the EU15 group.

Another effect can be seen here: a variable link between level of education and activity limitations. Depending on the country, the return on education varies, as does its relation to living conditions and determinants of health.

Figure 2: Prevalences of activity limitation and sizes of the 0-2 and 5-6 education groups, national prevalences and relative risks, by country for ages 30-79, observed and hypothetical.

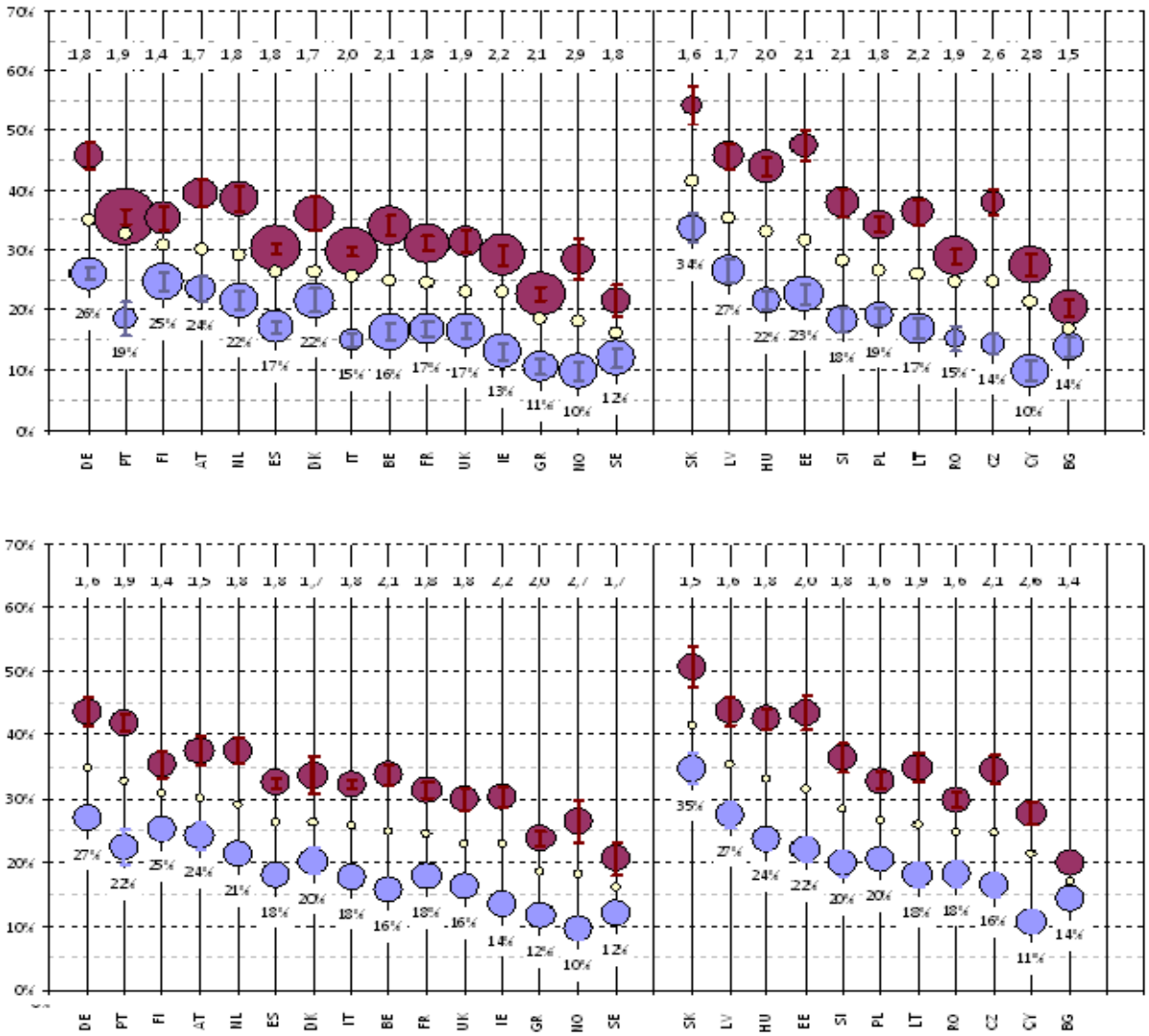
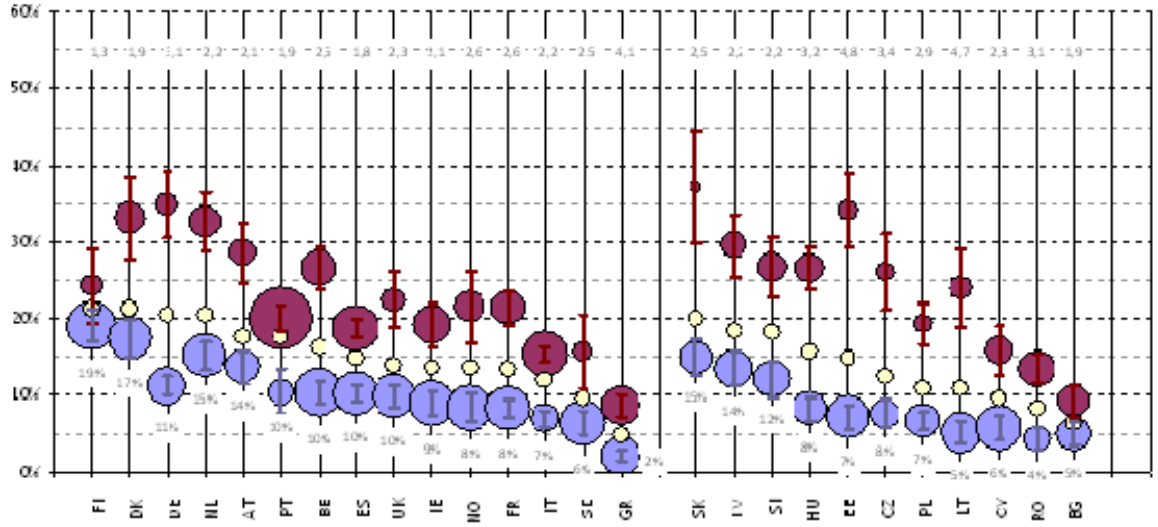
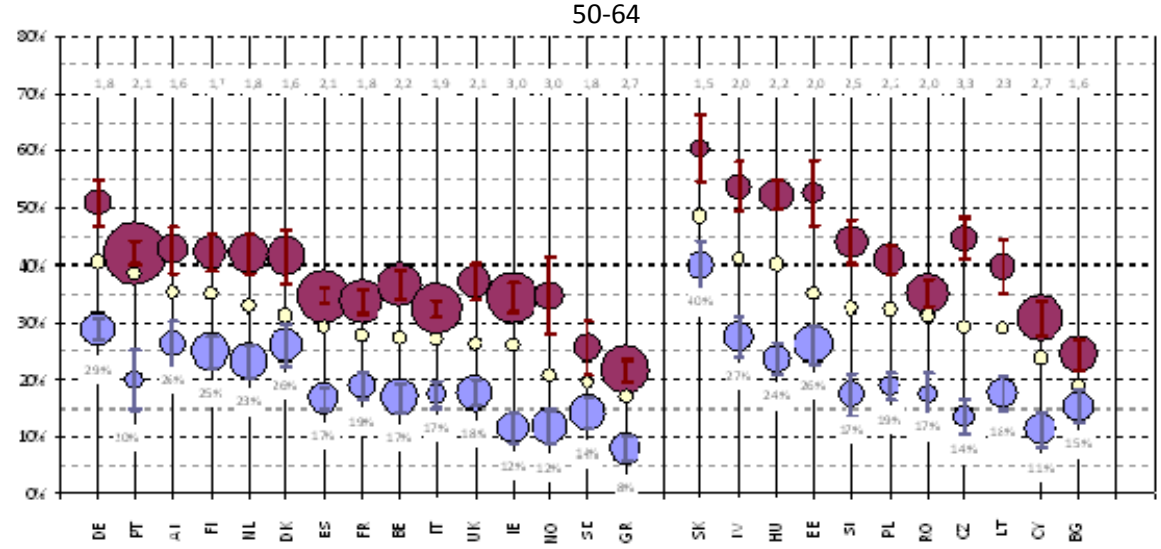


Figure 3: Prevalences and sizes of the 0-2 and 5-6 education groups, national prevalence and relative risks, by country and age group

30-49



50-64



65-79

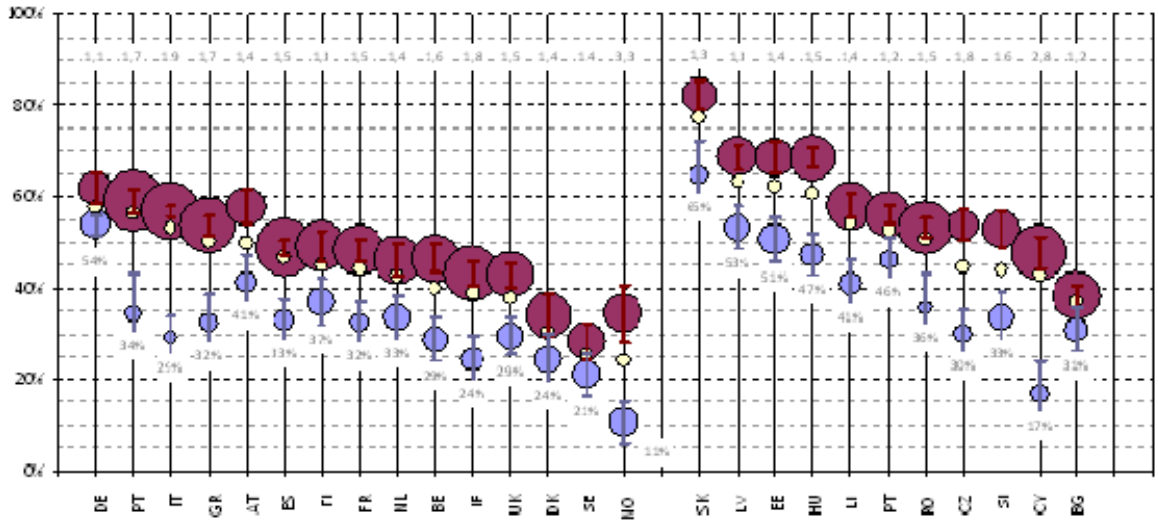
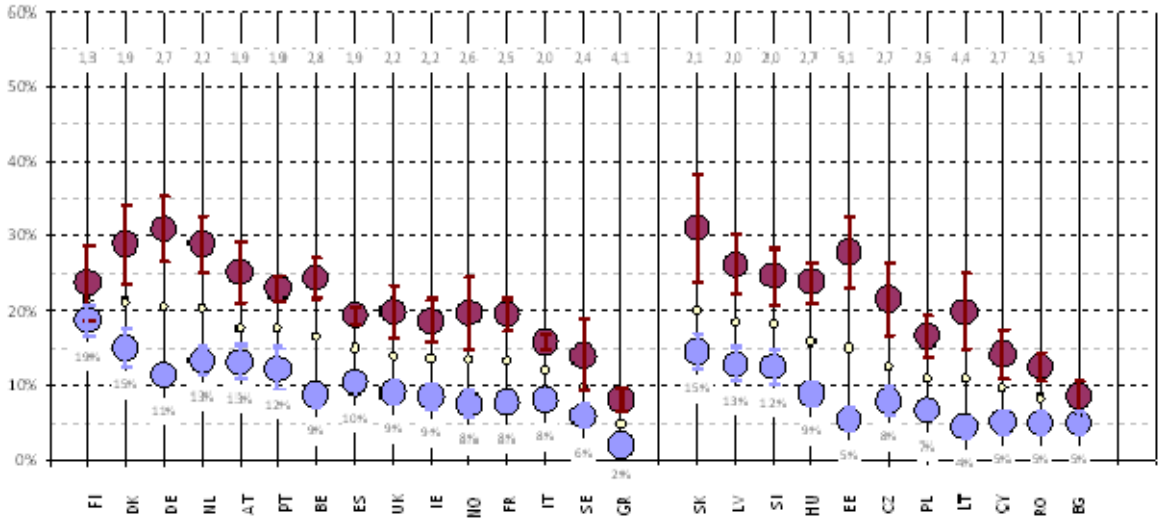
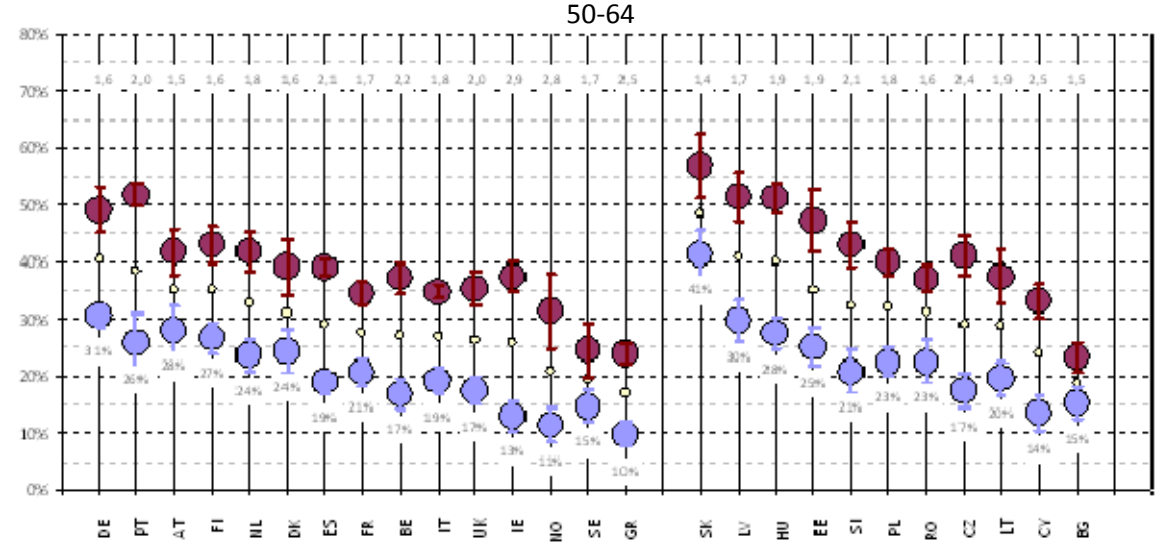


Figure 4: Adjusted or hypothetical prevalences of the 0-2 and 5-6 education groups (adjusted on a third of the population), national prevalence and relative risks, by country and age group

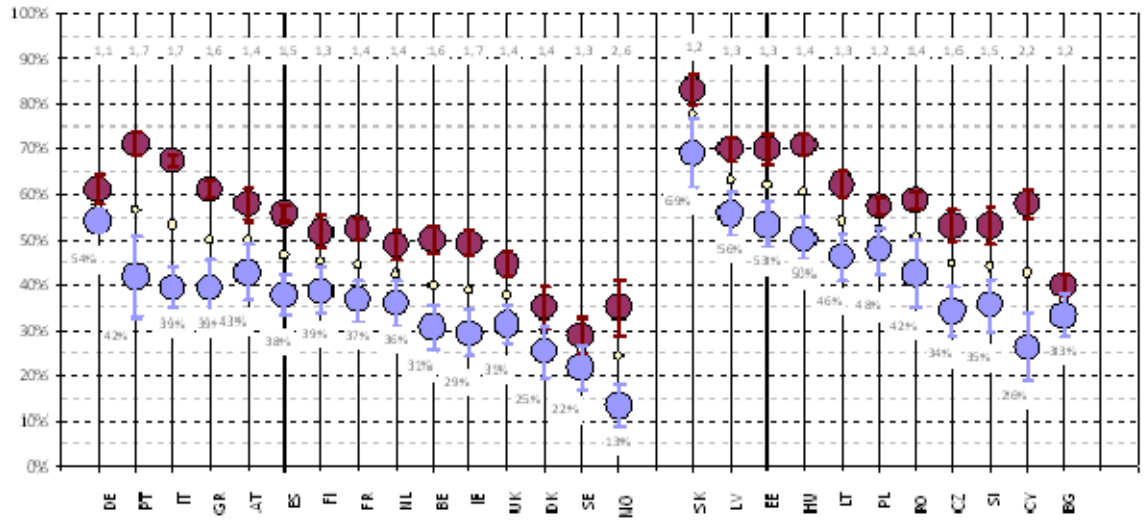
30-49



50-64



65-79



b. National prevalence of levels 5-6 as a function of relative risk

i. Countries' relative positions

We have also indicated the prevalences of the same three groups of cohorts and of the 5-6 level of education, as a function of hypothetical relative risks, for men and women in each country. We have also divided the countries into groups according to their date of entry into the European Union: the "first entrants" (AT, BE, DE, DK, ES, FI, FR, GR, IE, IT, NL, NO, PT, SE, UK) and those from the "fifth wave" of enlargement, mainly in the East (BG, CY, CZ, EE, HU, LT, LV, MT, PL, RO, SI, SK). The result is presented in Figure 4. The same graphs for 0-2 and 3-4 educational levels present identical patterns, but less marked differentials. This representation offers an overview of levels of inequality, as well as highlighting generational dynamics and disparities between educational levels and between eastern and western Europe. Men are represented by circles and women by diamonds; "western" countries by filled markers and "eastern" countries by hollow markers. On each graph, a cross centred on the centroid of the points highlights the relative positions and dynamics of the different points.

The higher the levels of education, the lower the centroid of the points, highlighting the lower prevalences of higher levels of education. The y-coordinate of the centroid remains fixed, since for each country, relative risk remains the same for different levels of education within a given generation. Going from the youngest cohorts to the oldest, the centroid moves upward and leftward. The vertical movement shows the increase in prevalence with age. The horizontal movement of the centroid (from 3.0 to near 1.5) shows a decrease in health disparities with age. This joint movement—the association between increasing age and both lower hypothetical relative risk and higher prevalence—is common to all countries, but it is greater in the eastern countries, with the exception of BG, CZ and SI, as well as DE and GR. The decrease in relative risk through the generations seems to be greater in men than in women.

Observation of the general shapes of the scatter plots seems to reveal an association between prevalence and relative risk, which is more or less marked depending on educational levels and cohort groups. Thus, for a given level of education and increasing age, there is a more marked decrease in prevalence as a function of hypothetical relative risk. For a given generation and with increasing educational level, besides a general decrease in prevalences, there is a clear tendency for the points to move closer to the regression line: the correlation coefficient, for men and women combined, lies between 0 and 0.1 for 0-2 educational levels, between 0.1 and 0.3 for levels 3-4 and between 0.3 and 0.5 for levels 5-6.

ii. Clustering using the k-means method

We would now like to establish a finer classification of groups of countries with similar models. This can be done using the k-means method.[3] This method is based on the identification of the groups which lie at the shortest distance from a given number of cluster centres. Here we identify 4 groups, on a plane formed by two axes: the prevalence of levels 5-6 and relative variation in prevalences

[3] Aravind H, C Rajgopal and K P Soman. Article:A Simple Approach to Clustering in Excel. *International Journal of Computer Applications* 11(7):19–25, December 2010. Published By Foundation of Computer Science.

between the 5-6 and 0-2 levels of education. These relative variations can be used to understand both health inequalities and relative risk. Each variable is normalized so as to prevent a given variable from disproportionately influencing another during clustering. Four groups were created for all individuals combined: ages 30-49, 50-64 and 65-79. We also performed a final 4-means clustering analysis with the same data for each of the three age groups of interest from each country. This analysis is performed not on a plane but in a six-dimensional space formed by the 5-6 prevalence axes for the 30-49, 50-64 and 65-79 age groups, and axes for the absolute variations in prevalence for the same age groups. This last clustering thus takes into account not only the countries' relative positions from the point of view of relative variations and the prevalences of high levels of education, but also the particular dynamics between generations in each country.

Figure 5 shows the result of country clustering for data from the entire 30-79 age range. The table that follows gives the results of all clustering analyses. For the clustering analyses on two axes, each group is numbered from 1 to 4: (1) corresponds to countries with high prevalences in the high-education group and low relative variations between the 5-6 and 0-2 education groups; 2 corresponds to countries with slightly lower prevalence and slightly larger variation between education groups; group 3 corresponds to countries with still lower prevalence in the high-education group and greater relative variation in prevalence; and finally group 4 represents countries with the lowest prevalence and highest relative variations. Finally, for the last clustering analysis, taking into account prevalences and their relative variations with education for each of the three age groups, the country groups are named from *a* to *d*, since they no longer correspond to relative positions on a simple trend line that can be represented on a two-dimensional graph. The results of this analysis can only be interpreted by observing the groups, or clusters, of countries in the three age groups.

In Figure 5 a tendency similar to the one observed in Figure 4 can be seen. Five countries are distinguished by low relative variations and high prevalence in the high-education group: AT, DE, FI, LV and SK (cluster 1); at the opposite extreme, three countries are distinguished by low prevalence of the 5-6 educational levels and high relative variations: CY, NO, CZ (cluster 4). The other countries fall in between these two extremes (clusters 2 and 3).

The last sets of clusters can be interpreted with the aid of the results of the clustering within each age group. Cluster *A* includes "middle-range" countries, situated, regardless of age group, in the intermediate clusters, 2 and 3 (with the exception of IE for the 50-64 age group, in position 4). It consists of BG, ES, FR, PT, BE, IE, IT, SE, SI and UK, with intermediate prevalences of educational levels 5-6 and intermediate levels of relative variations. The second group, B, which includes only CY and NO, is made up of countries systematically located in clusters 3 and 4, with low prevalences in the high-education group and high relative variations. Group *C* is made up of countries that systematically belong to clusters 1 and 2, with elevated prevalence in the high-education group and lower variation in prevalence in relation to the 0-2 group: it includes AT, DE, FI, LV, SK, DK and NL. Finally, group *D* includes countries with variably positioned age groups. Generally, in this cluster, increasing age is associated to a shift in position, from group 4 toward group 1; that is, from high relative variation and low prevalence in the high-education group toward high prevalence in the high-education group and low relative variation (except CZ). Countries from the "5th wave of integration" identified in Figure 4 are found in this group. Group *D* includes EE, HU, PL, GR, LT, RO and CZ.

Figure 4 : Prevalences of the 5-6 education groups as a function of hypothetical relative risk, by country, for ages 30-79

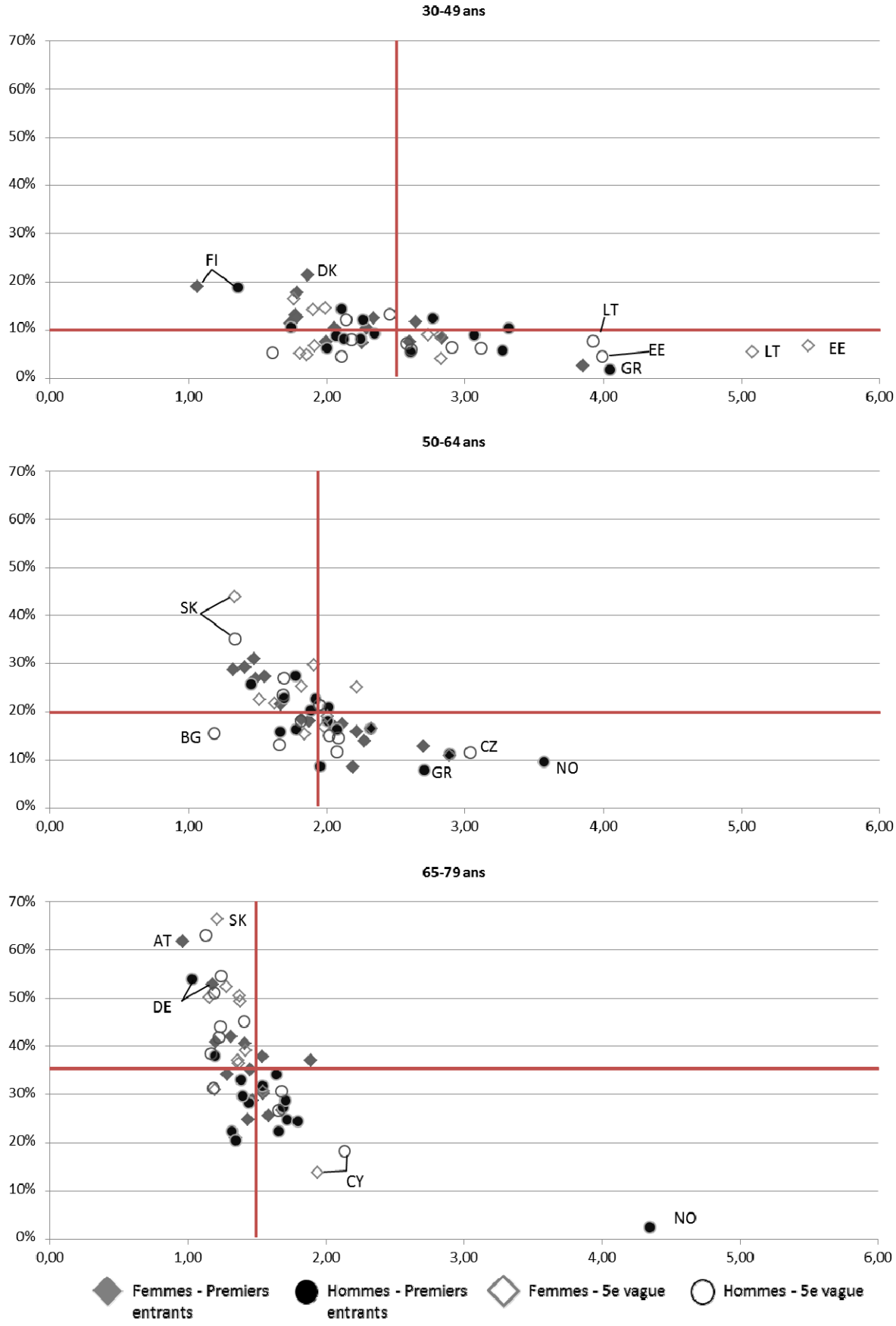


Figure 5: Four-means clustering of countries, according to relative variation in prevalence between educational levels 0-2 and 5-6 and the prevalence of the high-education group.

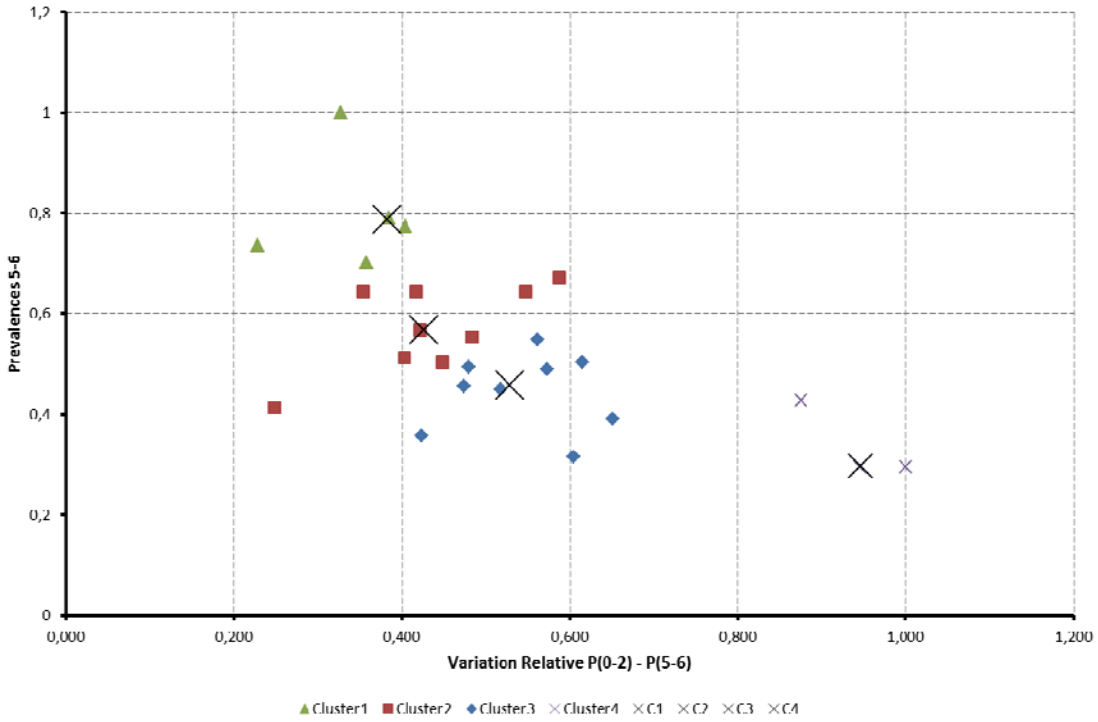


Table 3: Relative positions of the clustered countries, by age group, for all individuals, and taking into account all variables

	30-79	30-49	50-64	65-79	Cluster Total
BG	2	3	2	2	A
ES	2	2	3	2	A
FR	2	3	2	2	A
PT	2	2	3	3	A
BE	3	2	3	3	A
IE	3	2	4	3	A
IT	3	3	3	3	A
SE	3	3	3	2	A
SI	3	2	3	2	A
UK	3	2	3	2	A
CY	4	3	4	4	B
NO	4	3	4	4	B
AT	1	2	1	2	C
DE	1	2	1	1	C
FI	1	1	1	2	C
LV	1	2	2	1	C
SK	1	2	1	1	C
DK	2	1	1	2	C
NL	2	1	2	2	C
EE	2	4	2	1	D
HU	2	3	3	1	D
PL	2	3	3	1	D
GR	3	4	4	3	D
LT	3	4	3	2	D
RO	3	3	3	2	D
CZ	4	3	4	3	D

c. Logistic regression

The aim of this analysis is to verify, and if necessary complete, the initial results on health inequalities in Europe, based on activity limitations. In what follows (REF!) we present the results of the logistic regression calculating the probability of experiencing activity limitation (combining limitation and severe limitation). This regression controls for several factors: age, sex, level of education and country. We wish both to verify the initial results and also to identify specific effects in each country; thus, the interaction effect between education and country was also incorporated into the model.

The results show that age, naturally, has a significantly positive effect on the risk of experiencing activity limitation. Women seem to run a higher risk of activity limitation than men. The results on educational levels suggest that high levels of education limit risk, whereas levels 0-2 do not have a significant effect. This last result should be interpreted with caution, however, given the systematically positive effect of educational levels 0-2 at the country level, when significant. The suggestion of higher risk in certain countries with high national prevalence seems to be confirmed: SK, DE, LV, FI, EE and HU are the six countries where the probability that an individual will experience activity limitation is highest. PT is an exception, with relatively limited probabilities. At the opposite end, the initial suggestion of a lower risk of activity limitation in certain countries also seems to be confirmed. This is the case of NO and CY, which do not significantly differ from the reference SE, as well as GR, BG, IE and RO.

Finally, the analysis of education-country interaction effects confirms the existence of country-specific education effects, at least for a majority of 0-2 education groups. However, the lack of significance of the 5-6 educational levels in almost all countries makes it impossible to verify and refine observations on health inequalities at different educational levels by comparing the specific effects of a low level of education with those of a high level of education in each country.

Logistic regressions for the individual age groups were not significant enough to draw any conclusions on the existence of country-specific effects. Country risks, where significant, also confirm the results obtained using the bubble charts in Figures 3 and 4.

It is possible, however, to study possible avenues without taking non-significance into account, by looking at countries' rankings in both the logistic regressions and the bubble charts:

The two rankings are coherent with respect to national prevalences on the bubble charts and country probabilities by age group and for all individuals. Countries with high national prevalence also have a relatively high national probability of activity limitation. Only PT, which is among the countries with the highest national prevalences in the 50-64 and 65-79 age groups, is not among the countries with a high risk of activity limitation.

In a comparison not of national risks, but of the country-specific effects in the regressions with the prevalences from the bubble charts, larger differences appear:

- In the 30-49 age group: For educational levels 0-2, while CZ, GR and HU figure among the countries with low national prevalence, they also show marked specific effects, with a strong effect of this level of education in the population. At the opposite end, for FI, ES and PT,

which present relatively high national prevalence, the regression seems to show a very limited specific effect of low educational level.

Although BG's prevalences for all levels of education are apparently low, it seems here to present a high specific risk at high levels of education. In EE and DE, in contrast, specific effects are highly limited and prevalences high.

- In the 50-64 age group: At low levels of education, IE and NO show greatly increased risks of limitation, whereas their prevalence is relatively low, whereas in FI, AT and DE, while the prevalence of activity limitations is high, the risk revealed in the logistic regressions is low.

  At high levels of education, BG and IT show greatly increased risks and limited prevalence, whereas the results for HU, in contrast, suggest very limited specific risks.
- In the 65-79 age group: At education levels 0-2, NO is the only country with low prevalence of activity limitation but high increased risk, whereas in the high-prevalence countries LT, DE and PL, the logistic regression shows low specific risks.

For educational levels 5-6, no low-prevalence country seems to show high specific risks. However, SK, IT and LT present very limited specific effects and high prevalence.

Table 4: Probability of experiencing activity limitation by age, sex, level of education and country, taking into account the country/level of education interaction effect

Variable	Reference	Modality	Estimate	Variable	Reference	Modality	Estimate	Variable	Reference	Modality	Estimate
age			0.05 ***	Education x Country	3-4 / SE	0-2 GR	0.80 ***	Education x Country	3-4 / SE	5-6 DK	0.46 **
sex	M	W	0.11 ***			0-2 SK	0.69 ***			5-6 BG	0.29 *
Education	3-4	0-2	0,01			0-2 CZ	0.61 ***			5-6 AT	0,18
		5-6	-0.4 ***			0-2 EE	0.58 *			5-6 NL	0,12
Country	SE	SK	1.27 ***			0-2 HU	0.58 ***			5-6 PT	0,11
		DE	1.02 ***			0-2 IT	0.58 ***			5-6 ES	0,08
		LV	0.99 ***			0-2 PT	0.58 ***			5-6 IT	0,06
		FI	0.93 ***			0-2 CY	0.50 *			5-6 UK	0,06
		EE	0.91 ***			0-2 LT	0.50 **			5-6 FI	0,05
		HU	0.83 ***			0-2 LV	0.48 **			5-6 BE	0,04
		NL	0.69 ***			0-2 NO	0.46 **			5-6 FR	0,04
		AT	0.68 ***			0-2 BE	0.45 ***			5-6 SK	0,02
		SI	0.64 ***			0-2 PL	0.45 ***			5-6 DE	-0,0
		LT	0.50 ***			0-2 AT	0.44 ***			5-6 IE	-0,0
		PL	0.50 ***			0-2 IE	0.42 **			5-6 LV	-0,0
		ES	0.42 ***			0-2 RO	0.41 ***			5-6 PL	-0,0
		PT	0.42 ***			0-2 SI	0.40 *			5-6 SI	-0,0
		CZ	0.41 ***			0-2 DK	0.38 **			5-6 CY	-0,1
		DK	0.41 ***			0-2 FR	0.35 ***			5-6 EE	-0,1
		BE	0.36 ***			0-2 BG	0.34 **			5-6 GR	-0,1
		FR	0.35 ***			0-2 DE	0.34 ***			5-6 HU	-0,1
		UK	0.35 ***			0-2 ES	0.30 **			5-6 NO	-0,1
		RO	0.29 ***			0-2 NL	0.29 **			5-6 RO	-0,1
		IE	0.23 *			0-2 UK	0.29 **			5-6 CZ	-0,2
		IT	0.17 **			0-2 FI	-0,0			5-6 LT	-0,2
		CY	0,09								
		NO	0,05								
		BG	-0.1 *								
		GR	-0.3 ***								

30-49

Variable	Reference	Modality	Estimate	Variable	Reference	Modality	Estimate	Variable	Reference	Modality	Estimate
age			0.04 ***	Education x Country	3-4 / SE	0-2 CZ	0.65 *	Education x Country	3-4 / SE	5-6 DK	0.56 **
sex	M	W	0.11 ***			0-2 GR	0.55 *			5-6 AT	0.45 *
Education	3-4	0-2	0,28			0-2 SK	0,49			5-6 BG	0,45
		5-6	-0.6 ***			0-2 DK	0.47 *			5-6 FI	0.41 *
Country	SE	FI	0.84 ***			0-2 EE	0,47			5-6 NL	0.40 *
		DE	0.77 ***			0-2 HU	0.43 *			5-6 PT	0,31
		SK	0.71 ***			0-2 NL	0.42 *			5-6 LV	0,26
		NL	0.60 ***			0-2 AT	0.41 *			5-6 SK	0,24
		DK	0.59 ***			0-2 RO	0.38 *			5-6 ES	0,23
		LV	0.57 ***			0-2 DE	0.37 *			5-6 UK	0,22
		SI	0.54 ***			0-2 PL	0,37			5-6 IE	0,20
		EE	0.49 *			0-2 LT	0,36			5-6 SI	0,20
		AT	0.43 ***			0-2 LV	0,33			5-6 CZ	0,18
		BE	0.43 ***			0-2 BE	0,28			5-6 IT	0,18
		ES	0.33 ***			0-2 FR	0,27			5-6 FR	0,16
		HU	0.32 **			0-2 CY	0,22			5-6 BE	0,11
		UK	0.29 **			0-2 SI	0,22			5-6 PL	0,09
		IE	0.27 *			0-2 UK	0,19			5-6 CY	0,08
		NO	0.27 *			0-2 BG	0,18			5-6 NO	0,06
		PT	0,22			0-2 NO	0,16			5-6 RO	0,01
		FR	0.16 *			0-2 PT	0,15			5-6 GR	-0,0
		LT	0,07			0-2 IT	0,14			5-6 HU	-0,0
		CZ	0,06			0-2 ES	-0,0			5-6 DE	-0,1
		IT	-0,0			0-2 IE	-0,0			5-6 EE	-0,3
		PL	-0,0			0-2 FI	-0,2			5-6 LT	-0,3
		CY	-0,1								
		RO	-0.5 ***								
		BG	-0.6 ***								
		GR	-1.1 ***								

50-64

Variable	Reference	Modality	Estimate	Variable	Reference	Modality	Estimate	Variable	Reference	Modality	Estimate
age			0.05 ***	Education x Country	3-4 / SE	0-2 IE	0.70 **	Education x Country	3-4 / SE	5-6 DK	0.53 *
sex	M	W	0.09 ***			0-2 PT	0.62 *			5-6 BG	0,36
Education	3-4	0-2	0,11			0-2 DK	0.60 **			5-6 IT	0.29 *
		5-6	-0.5 ***			0-2 NO	0.57 *			5-6 ES	0,14
Country	SE	SK	1.27 ***			0-2 CZ	0.56 **			5-6 FR	0,14
		LV	1.00 ***			0-2 BE	0.51 **			5-6 SK	0,14
		DE	0.99 ***			0-2 IT	0.49 **			5-6 BE	0,12
		HU	0.90 ***			0-2 GR	0.45 *			5-6 PT	0,11
		FI	0.87 ***			0-2 EE	0,44			5-6 AT	0,05
		EE	0.80 ***			0-2 ES	0.44 **			5-6 NL	0,05
		AT	0.69 ***			0-2 UK	0.44 **			5-6 UK	0,05
		NL	0.58 ***			0-2 LT	0,41			5-6 DE	-0,0
		SI	0.56 ***			0-2 SI	0,41			5-6 EE	-0,0
		PL	0.54 ***			0-2 CY	0,39			5-6 IE	-0,0
		RO	0.51 ***			0-2 HU	0.37 *			5-6 CY	-0,1
		LT	0.43 **			0-2 BG	0,33			5-6 FI	-0,1
		CZ	0.38 ***			0-2 LV	0,33			5-6 LV	-0,1
		PT	0,26			0-2 SK	0,31			5-6 NO	-0,1
		DK	0.25 *			0-2 NL	0.29 *			5-6 PL	-0,1
		FR	0.22 **			0-2 PL	0.28 *			5-6 GR	-0,2
		UK	0.22 **			0-2 FR	0.26 *			5-6 HU	-0,2
		BE	0,12			0-2 DE	0,23			5-6 LT	-0,2
		ES	0,11			0-2 AT	0,20			5-6 RO	-0,2
		CY	0,00			0-2 RO	0,06			5-6 SI	-0,2
		IT	-0,0			0-2 FI	-0,0			5-6 CZ	-0.4 *
		NO	-0,0								
		IE	-0,1								
		BG	-0.2 *								
		GR	-0.5 ***								

65-79

Variable	Reference	Modality	Estimate	Variable	Reference	Modality	Estimate	Variable	Reference	Modality	Estimate
age			0.06 ***	Education x Country	3-4 / SE	0-2 PT	0,53	Education x Country	3-4 / SE	5-6 DE	0,05
sex	M	W	0.16 ***			0-2 IT	0.38 *			5-6 PL	0,03
Education	3-4	0-2	0,17			0-2 CY	0,37			5-6 DK	0,01
		5-6	-0,1			0-2 HU	0.33 *			5-6 AT	0,00
Country	SE	SK	2.35 ***			0-2 BE	0,30			5-6 BE	-0,0
		EE	1.68 ***			0-2 NO	0,30			5-6 BG	-0,0
		LV	1.62 ***			0-2 GR	0,27			5-6 ES	-0,0
		DE	1.40 ***			0-2 AT	0,25			5-6 FI	-0,0
		HU	1.37 ***			0-2 CZ	0,21			5-6 NL	-0,0
		LT	1.28 ***			0-2 IE	0,19			5-6 PT	-0,0
		PL	1.10 ***			0-2 ES	0,16			5-6 UK	-0,0
		RO	1.04 ***			0-2 SI	0,16			5-6 FR	-0,1
		AT	0.98 ***			0-2 SK	0,12			5-6 SI	-0,1
		IT	0.84 ***			0-2 LV	0,11			5-6 GR	-0,2
		CZ	0.83 ***			0-2 DK	0,10			5-6 HU	-0,2
		GR	0.81 ***			0-2 FR	0,10			5-6 LV	-0,2
		PT	0.79 *			0-2 UK	0,10			5-6 RO	-0,2
		SI	0.79 ***			0-2 EE	0,09			5-6 CY	-0,3
		FI	0.78 ***			0-2 FI	0,09			5-6 CZ	-0,3
		NL	0.77 ***			0-2 PL	0,06			5-6 EE	-0,3
		FR	0.73 ***			0-2 DE	-0,0			5-6 IE	-0,3
		ES	0.72 ***			0-2 LT	-0,0			5-6 LT	-0,3
		BG	0.58 ***			0-2 NL	-0,0			5-6 IT	-0.4 *
		UK	0.50 ***			0-2 RO	0,00			5-6 SK	-0,4
		CY	0,47			0-2 BG	-0,1			5-6 NO	-0.7 *
		IE	0.47 *								
		BE	0.46 **								
		DK	0,16								
		NO	-0,1								

Annex: Educational classification

Level	English	Principal characteristics[4]
0	Pre-primary level	*Introduces very young children to a school-type environment, i.e. provides a bridge between the home and a school-based atmosphere.*
1	Primary level	*Gives students a sound basic education in reading, writing and mathematics along with an elementary understanding of other subjects such as history, geography, natural science, social science, art and music. In some cases religious instruction is featured. Programmes normally designed on a unit or project basis trather than by subjects.*
2	Lower secondary level	*Completes the provision of basic education which began at the preceding level, with the aim of laying the foundation for lifelong learning and human development. The programmes at this level are usually on a more subject-oriented pattern using more specialized teachers (which differentiates it from level 1).*
3	Upper secondary level	*Typically begins at the end of full-time compulsory education for those countries that have a system of compulsory education. Often teachers need to be more qualified or specialized than at the preceding level.* *The minimum entrance requirements are the completion of level 2 or demonstrable ability to handle programmes at this level.*
4	Post-secondary non tertiary level	*Programmes that straddle the boundary between levels 3 and 5 from an international point of view. Prepares students for studies at level 5 who, although having completed ISCED level 3, did not follow a curriculum which would allow entry to level 5. Includes short vocational programmes which require completion of level 3 for entry.*
5	First stage of tertiary level	*Educational content more advanced than those offered at levels 3 and 4, in programmes with a cumulative theoretical duration of at least 2 years from the beginning of level 5.* *Does not lead directly to the award of an advanced research qualification (level 6).*
6	Second stage of tertiary level	*Reserved for tertiary programmes which lead to the award of an advanced research qualification. The programmes are therefore devoted to advanced study and original research and are not based on course-work only.*

[4] http://www.unesco.org/education/information/nfsunesco/doc/isced_1997.htm, consulted on 13 Feb 2014

Co-funded by 10 Member States, the European Commission, DG SANCO and two French institutions: DREES and CNSA.

Contact EHLEIS:
Jean Marie ROBINE, INSERM
Université Montpellier II / U710 - MMDN
Place Eugène Bataillon, bat 24 - CC105
34095 Montpellier Cedex 05, France.

Tel: +33 (0) 467 14 33 85
Fax: +33 (0) 467 14 92 95
Email: jean-marie.robine@inserm.fr

www.eurohex.eu