# Literature Review

**October 2009**

Work package 6 – Typology of regions

Work package 7 – Typology of infra regional level

October 2009

# Background

The I2SARE project is about the collection and efficient use of health inequalities indicators in the regions of Europe. The project aims to assist European, national, regional and local decision makers in developing their health policy through a better understanding of the heath status of their population and of health inequalities at regional and sub regional level.

In work package 6 and 7 a typology of European regions and sub-regions is to be developed according to demographic, socio-economic and health related indicators. By doing so regions with similarities throughout Europe are described and territories are grouped according to the comparability of health determinants. This will enable regions to select appropriate regions for benchmarking analyses and for comparisons and cooperation projects in the future.

The typology of regions and sub-regions will be performed by a cluster analysis, which is a statistical methodology to group subjects (e.g. regions) according to predefined variables in unknown, homogenous clusters which are heterogeneous between each other. Before this analysis is performed a literature review is to be performed to identify variables which will be included in the cluster analysis and will clarify the appropriateness of this methodology. The first part of the literature search will predominately identify variables which are important as determinants of the health status of a population and as determinants for the success of health programmes and policies.

This literature review will be performed for the regional and sub-regional level. The results of both reviews are presented in this report.

# Research Questions - WP6

1. Which health indicators are important as determinants for the health status of a population?
2. Which determinants describe the success of health programmes and policies?
3. What examples exist for cluster analysis at a regional level and what methodology is used?

# Methods – WP6

In January 2009 a literature search was performed on the afore mentioned research questions. A search was carried out in English and German publications without restrictions on time. The type of publications covered scientific publications as well as grey literature and government white papers.

Whilst the literature search on the first two research questions was guided by the experiences and resources from other projects, especially by the two main European indicator projects ECHI 1 and 2 as well as ECHIM and ISARE 1 to 3, the literature search on the third research question was performed by a systematic search strategy in the Medline database. The search terms used were “cluster analysis, classification, benchmark, comparison, population, group”, the search strategy can be found in annex 1.

Furthermore the google search engine and websites from important organisations such as WHO, OECD and Eurostat were used, to identify literature on all three research questions.

Regarding the third research question on examples and methodologies to develop a typology of regions, only literature was included which had populations’ health and the clustering of geographic areas under focus.

# Results – WP6

## *Literature search health indicators*

Areas of relevant indicators to determine the health of the population are defined in various fundamental publications. The Lalonde report from 1974 should be mentioned, which defines a public health model with four main categories that determine health, i.e. biological and genetic factors, lifestyle, the environment and the health care system [1]. Moreover, Göran Dahlgren and Margret Whitehead have defined the social determinants of health as age, sex and constitutional factors, individual life style factors, social and community networks, living and working conditions as well as general socio-economic, cultural and environmental conditions [2].

Based on this background, sets of indicators relevant to determine the health status of the population on a European level have been defined by the European funded project "European Community Health Indicators (ECHI 1 + 2)" and are updated and specified by the follow-up project ECHIM (European Community Health Indicator Monitoring). The indicators assembled by the ECHI projects cover the following categories

- Demographic and socio-economic factors
- Health status
- Determinants of health
- Health services
- Health promotion

The latest published ECHI shortlist covers 88 indicators which are important for the assessment of the overall health status and major health problems at population level. The indicators have been chosen by public health generalists according to the strength of evidence in measuring inequalities in health and according to their importance for assessing effective interventions and health policies [3]. A detailed list of shortlist indicators as well as a metadata description can be found on www.healthindicators.org.

According to the ECHI short list of relevant health indicators the predecessor projects of I2SARE have assembled a list of relevant and available indicators on regional level. The list contains 75 indicators covering the topics [4]

- Health professionals
- Health care services
- Demographic and socio-economic data
- Mortality data
- Morbidity data
- Risk factors
- Data on prevention

Despite 28 indicators all indicators identified by the ISARE 3 project are also included in the latest ECHI short list. Mainly the ISARE 3 indicators which are not represented in the ECHI short list are absolute numbers for calculating an indicator.

It was also asked which kind of indicators can be used to display socio-economic inequalities in health or to assess the socio-economic status of a population. Regarding this question publications from the regional office for Europe of the World Health Organisation are identified [5, 6]. In the report "Measuring socio-economic inequalities in health" and successive publications the most important indicators are introduced which are occupational status, level of education and the income level. The report concluded if only one indicator could be selected the measurement of the educational level in the population would be recommended as the indicator of choice[5].

## Literature search on health performance measures

A recently published background report for the WHO European Ministerial Conference on Health Systems "Health Systems, Health and Wealth" describes the requirements for health performance measures [7].

It defines health performance indicators as measurements that "seek to monitor, evaluate and communicate the extent to which various aspects of the health system meet their key objectives" [8]. There is no useful single method of reporting health performance as there are different needs of various stakeholder groups (e.g. government, tax payers, physicians, patients). Possibilities for health performance measurements could be at the population level, the individual level, or measurements on the responsiveness of the health system or the measurement of the productivity of the system could be included [8]. As the indicators in I2SARE are limited only a relatively small number of data items can be included in the development of a typology [8].

## Literature search on examples and suitable methods for the development of a typology of regions

The literature-search in Medline resulted in 464 publications but after scanning the title and/or abstract only two publications were relevant to clustering geographical regions for health comparison purposes [9, 10]. Also, there was some literature identified by google search engine [11-13] and from suggestions of literature from members of the institute [14, 15].

Most of the studies used a K-means clustering technique [9, 10, 12-15]. In one publication the clustering technique was not specified. The K-means clustering is a non-hierarchical clustering technique which predefines the desired number of clusters. All objects are assigned to one of the predefined clusters by defined selection or by an arbitrary procedure. According to the distance of each object from the centroid of the clusters the cases are then reassigned until each object is assigned to a cluster with its closest distance to the respective centroid [16]. Some publications did a hierarchical cluster analysis first where at the outset each object builds a cluster and by adding together objects with the smallest distance measures, the clusters are build until in the end there is only one cluster left. After the number of clusters are estimated the analysis was repeated and validated by a K-means clustering technique [14, 15].

The variables used to define the clusters varied regarding to the respective research questions. The variables or the main categories can be found in the summary table 1. Two studies performed a factor analysis to minimise the variable set for the cluster analysis [13, 15].

**Table 1** Publications of studies which used cluster analysis for geographical clustering for health comparison purposes

| Author | Clustering technique | Variables included | Research question / Research objective |
|---|---|---|---|
| Ruger JP et al. [9] | K-means | < 5 year total mortality<br><br>Adult mortality (men, women) | Development mortality strata of countries worldwide regarding the mortality rate in children (<5 years) and adult mortality rate (better-off, worse-off, mid-level) for further comparisons. |
| Siri JP et al. [10] | K-means | (1) household access to electricity and to piped water, (2) ownership of the dwelling, (3)education level of the primary caregiver, (4) distance from the city centre, (5) population density, (6) normalized difference vegetation index. | To identify homogenous urban units in the model region Kisumu, Kenya according to variables which are known to be related to malaria occurrence for the identification of areas where malaria interventions are needed most. |
| Niederländer E [11] | Not known | Total mortality rate in 5-year age groups | To identify groups of European countries according to their national mortality profile by age |
| McNabb L [12] | K-means | 24 variables measuring (1) Population Change, (2) Demographic Structure, (3) Social Status, (4) Economic Status, (6) Ethnicity, (7)<br><br>Aboriginal Status, (8) Housing, (9) Urbanisation/Metropolitan Influence, (10) Income Inequality, (11) Labour Market Conditions | To identify comparable regions according to the distribution of socio-economic variables. |
| Odoi A et al. [13] | K-means | 18 variables measuring (1) demographic structure, (2) social status, (3) economic status, (4) ethnicity, (5) aboriginal status, (6) housing<br><br>After a factor analysis 5 factors are included in the cluster analysis. | To identify and classify socio-economic and demographic characteristics of neighbourhoods in the city of Hamilton, Ontario, Canada. |

| | | | |
|---|---|---|---|
| Escamilla Loredo MI [14] | First:Hierarchical cluster analysis. Afterwards: K-means | (1) Accumulation annual growth rate of the population, (2) population density, (3) proportion of 1-year population in relation to total population, (4) population aged >65 in relation to the total population, (5) sex ratio, (6) proportion of females aged 50-65 in relation to total female population, (7) unemployment rate, (8) Disposable income of household per inhabitant, (9) Number of physicians per 100,000 inhabitants, (10) Type of health care system, (11) Type of government system | To identify groups of comparable European regions to allow benchmarking and the identification of good practice health models. |
| Strohmeier KP et al.[15] | First: Hierarchical cluster analysis. Afterwards: K-means K-means | (1) Percentage 0-14 years old in the population, (2) percentage >65 in the population, (3) Percentage of migrants, (4) population density, (5) population change in the last 5 years, (6) recipients of social aid/1000, (7) Unemployment, (8) disposable income<br><br>After a factor analysis 2 factors are included in the cluster analysis. | To identify regions with comparable socio-economic patterns in North Rhine-Westphalia for comparisons of health-related variables. |
| Trugeon A et al.[20] | First: Hierarchical cluster analysis. Afterwards: K-means | 50 variables measuring (1) Demographic Structure, (2) Social Status, (3) Economic Status, (4) Housing, (6) Elderly care, (7) Care offer, (8) Mortality | Observation of the social and health inequalities in France. |

# Discussion – WP6

The literature review shows that there are already lists of recommended health determinants available. As it is one of the aims of the European Commission to use and produce comparable health information, the indicators recommended by previous indicator projects such as ECHI 1+2 and ECIM as well as ISARE 1-3 should be used when developing a typology on regional or sub-regional level. Life expectancy and infant mortality might be the only available indicators in I2SARE on the topic health performance measurements.

The identified literature on cluster analysis shows that, despite the publication by Siri and colleagues [10], comparable indicator sets to those defined by ECHIM are used. Sometimes the cluster analyses focussed just on one aspect such as mortality rates in the respective countries [9, 11] or on socio-economic patterns of the geographical area [12, 13, 15] but also cluster analyses including a wider range of indicators have been found [14]. When starting the analysis, testing to see if factor analysis, to reduce the number of dependent variables should be performed especially if several strong correlations amongst the indicators could exist. Furthermore, as it is not known in how many clusters the approximately 250 European regions might be separated into, a hierarchical clustering technique should be done first and afterwards this should be verified by a K-means cluster analysis.

# The sub regional level case – WP7

In the case of the WP7, there are two particular difficulties. First, due to the subregional character, we are classically confronted with poor quality data. Indeed, the main difficulty encounter in subregional level analysis is the lack of cases to run traditional statistics based on frequency studies. Secondly, due to the subregional character at a European scale, some data are simply unavailable for some subregional areas.

Generally, the few analyses which can be found in recent publications about clustering at a subregional level are focused on the purpose to determine geographical clusters through non-physical borders and about a specific disease [17-19]. There are mainly two reasons for this focus. The first reason is the will desire to change local health policy: the use of administrative geographical definition of territory to provide health analysis is often unsatisfactory. The other reason is that Bayesian analysis has allowed statistical work with a very low number of cases for the past 10/15 years.

The study published by FNORS attempts to suggest an analysis at a subregional level based on 50 health and socioeconomic indicators to cluster cantons in France, with no consideration for the geographic proximity, and compare groups of cantons. The purpose of the clustering was to group cantons by profile with similar profiles, even if those cantons are at a considerable geographical distance from each other. The applied methodology is base on a principal component analysis and a hierarchical cluster analysis, followed by a K-means consolidation. Finally, the 3.168 cantons are distributed in eight groups, corresponding with height clear socioeconomics and health profiles [20].

# Literature


1. Lalonde M. A new perspective on health of the Canadians. Ottawa: Ministry of National Wealth and Welfare; 1974

2. Dahlgren G, Whitehead M. Policies and Strategies to promote Equity in Health. Copenhagen: WHO, Regional Office Europe; 1992

3. Kipeläinen K, Aroma A, and the ECHIM Core Group. European Health Indicators: Development and Initial Implementation. Helsinki: National Public Health Institute; 2008

4. ISARE 3. Health Indicators in the European Regions - Final Report ISARE 3. Paris: Fédération nationale des observatoires régionaux de la santé; 2007

5. Kunst A, Mackenbach JP. Measuring socioeconomic inequalities in health. Copenhagen: WHO, Regional Office Europe; 2000

6. Commission on Social Determinants on Health. A Conceptual Framework for Action on the Social Determinants of Health. Discussion paper for the Commission on Social Determinants of Health DRAFT. Geneva: WHO; 2007

7. Smith PC, Mossialos E, Papanicolas E. Performance measurement for health system improvement: experiences, challenges and prospects. Copenhagen: WHO Europe; 2008

8. Smith PC, Mossialos E, Papanicolas I. Performance measurement for health system improvement: experiences, challenges and prospects. Copenhagen: WHO and WHO on behalf of the European Observatory on Health Systems and Policies; 2008

9. Ruger JP, Kim HJ. Global health inequalities: an international comparison. J Epidemiol Community Health. 2006 Nov;60(11):928-36.

10. Siri JG, Lindblade KA, Rosen DH, Onyango B, Vulule J, Slutsker L, et al. Quantitative urban classification for malaria epidemiology in sub-Saharan Africa. Malar J. 2008;7:34.

11. Niederlaender E. Causes of death in the EU - Marked differences between old and new EU member states: Eurostat; 2006

12. MacNabb L. Health Region Peer Groups. Statistics Canada; 2002. p. 85-90.

13. Odoi A, Wray R, Emo M, Birch S, Hutchison B, Eyles J, et al. Inequalities in neighbourhood socioeconomic characteristics: potential evidence-base for neighbourhood health planning. Int J Health Geogr. 2005 Aug 10;4:20.

14. Maria Ixhel Escamilla Loredo. Clustering of European regions as basis for benchmarking health programmes and interventions. Bielelfeld: University of Bielefeld; 2007.

15. Strohmeier KP, Schultz A, Bardehle D, Annuss R, Lenz A. [Health indicator-based cluster analysis of districts and urban districts in North Rhine-Westphalia]. Gesundheitswesen. 2007 Jan;69(1):26-33.

16. Afifi AA, V C. Computer-aided multivariate Analysis. 2 ed. New York: Van Nostrand Reinhold Company; 1990.

17. Bilancia M, Fedespina A. Geographical clustering of lung cancer in the province of Lecce, Italy: 1992-2001. Int J Health Geogr. 2009;8:40.

18. Borden KA, Cutter SL. Spatial patterns of natural hazards mortality in the United States. Int J Health Geogr. 2008;7:64.

19. Sugumaran R, Larson SR, Degroote JP. Spatio-temporal cluster analysis of county-based human West Nile virus incidence in the continental United States. Int J Health Geogr. 2009;8:43.

20. Trugeon A, Fontaine D, Lémery B. Inégalités socio-sanitaires en France. De la région au canton. Paris: MASSON; 2006

# Annex 1

Search strategy Medline (January 2009)

| Search | Most Recent Queries | Results |
|---|---|---|
| **#13** | **Search #11 NOT #12 Limits: Humans, Comparative Study, Journal Article, English, German** | **464** |
| #12 | Search **gene*[Title/Abstract]** Limits: **Humans, Comparative Study, Journal Article, English, German** | 375040 |
| #11 | Search **#4 AND #9 AND #10** Limits: **Humans, Comparative Study, Journal Article, English, German** | 520 |
| #10 | Search **#5 OR #7** Limits: **Humans, Comparative Study, Journal Article, English, German** | 521971 |
| #9 | Search **#6 OR #8** Limits: **Humans, Comparative Study, Journal Article, English, German** | 1361021 |
| #8 | Search **population OR group[Title/Abstract]** Limits: **Humans, Comparative Study, Journal Article, English, German** | 1164010 |
| #7 | Search **comparison[Title/Abstract]** Limits: **Humans, Comparative Study, Journal Article, English, German** | 286978 |
| #6 | Search **country OR region[Title/Abstract]** Limits: **Humans, Comparative Study, Journal Article, English, German** | 248251 |
| #5 | Search **classification OR benchmark[Title/Abstract]** Limits: **Humans, Comparative Study, Journal Article, English, German** | 249309 |
| #4 | Search **cluster analysis[Title/Abstract]** Limits: **Humans, Comparative Study, Journal Article, English, German** | 3846 |