A European Health and Environment Information System for Exposure and Disease Mapping and Risk Assessment

EUROHEIS

FINAL REPORT 2002

Project Code SI2.291820 (2000 CVG2-605)

SAHSU, Dept. Epidemiology & Public Health, Imperial College	UK
National Board of Health	Denmark
National Public Health Institute	Finland
Trinity College, University of Dublin	Ireland
WHO European Centre for Environment & Health, Rome	Italy
Dept. Statistics & Operation Research, University of Valencia	Spain
Dept. Epidemiology, Stockholm Centre of Public Health	Sweden

6 August 2002

CONTENTS

EXECUTIVE SUMMARY

The EUROHEIS (European Health and Environment Information System) project aims to improve understanding of the links between environmental exposures, health outcome and risk through the development of integrated information systems for rapid assessment of relationships between the environment and health at a geographical level. EUROHEIS is a three-year project with each of three phases funded separately. The first phase was feasibility study, performed in the first year of the project (15 December 1999 - 14 December 2000; previously reported as the final report for Phase I). The present report concerns Phase II. The initial project period was from 15 December 2000 until the 15 December 2001, however this was subsequently extended to end on 15 May 2002. Although this report is for Phase II, it should be regarded as an interim report for the entire project.

Phase I assessed the possibilities of implementing systems for point source investigations and disease and exposure mapping within the participating countries, modelled on a system (the Rapid Inquiry Facility (RIF)) being developed within the UK. Subsequent to the feasibility study, partners were identified for full or partial implementation of the system. The UK RIF has so far been installed in Spain and Sweden with some modifications. In particular, modules have been added by the Spanish partner enhancing the system's capabilities.

In Finland the existing SMASH system has been modified in collaboration with the EUROHEIS group. The RIF system is also being implemented and a comparison between the two systems will be made in Phase III. Denmark has decided to adapt the RIF to their own requirements, which will take advantage of their rich data sources incorporating individual level population data.

Ireland undertook an extensive review of meta data protocols for the exchange of meta data between partners in relation to the health and socio-economic databases available to the partners in order to develop and test appropriate and sensitive measures of deprivation at a suitable spatial level. Italy have explored the possibility of using the RIF for health impact assessment. The Italian partner has also been responsible for liaising with APHEIS (an Air Pollution and Health Information System), which is another project funded by the same EU programme as EUROHEIS (DG SANCO, Action on Pollution Related Research)

The UK partner has been further developing the Rapid Inquiry Facility (RIF) to aid its implementation in several of the partner countries. The scripts have been rewritten in light of feedback from our partners. The system has been tested and finalised ready for roll-out. Documentation, including the data and hardware specifications has been completed and we have provided assistance (including site visits) to partner countries where required. The UK partner has taken the lead on the development of the EUROHEIS website which includes details of the project, project partners, interim reports, and a demonstration of the RIF as well as information on the end of project conference planned for 30-31 March 2003. A leaflet summarising the EUROHEIS project has also been produced. The results of the project have been disseminated through international meetings, publications and through the project website.

During Phase II the RIF has been successfully implemented at various levels within the partner countries. In some countries the facility will be further developed and enhanced during Phase III. The public health benefits of the system will be demonstrated in Phase III through a series of case studies designed in Phase II.

INTRODUCTION

Background

The project aims to improve understanding of the links between environmental exposures, health outcome and risk through the development of integrated information systems for rapid assessment of relationships between the environment and health at a geographical level. The partners involved in this project have been among the leaders in Europe in developing and applying epidemiological and statistical methods for the appraisal and analyses of disease occurrence in small geographical areas related to point source exposures and to disease mapping.

EUROHEIS is a three-year project with each of three phases funded separately. This is the report for Phase II. The initial project period was from the 15 December 2000 until the 15 December 2001, however this was subsequently extended to end on the 15 May 2002. Although this report is for Phase II, it is in fact an interim report for the entire project.

In the first phase of the project, a one-year feasibility study was undertaken to assess the possibilities of implementing systems for point source investigations and disease and exposure mapping within the participating countries, modelled on a system being developed within the UK. The UK system was also further developed to include more generalised modules for disease and exposure mapping. It suggested that data are available and suitable for the development of a system for point-source analysis and disease mapping in several of the partner countries. Prototype systems for point-source investigations, including some disease mapping facilities, had already been developed in the UK and Finland.

This report concerns the work of the second phase of the project, building on the feasibility study. Its aim was to implement systems for point-source investigations and disease and exposure mapping (incorporating state-of-the-art Bayesian statistical models and Geographical Information Systems (GIS)) in the participating countries, modelled on the system being developed by the UK partner.

Objectives

The following work packages were identified and formed the specific objectives for the phase II.

Work package	Objective
1	Data Collection (January to June 2001) Health event data, population denominators as well as data on environmental exposures and other major explanatory variables (including socio-economic indicators) will be collected as appropriate. This is a major effort in some of the partner countries, whereas some countries already have most of these data available.
2	Development of computerised systems (January to December 2001) Depending on the outcome of the feasibility study, computerised systems will be developed in the partner countries, reflecting the different levels of data detail/ aggregation. These systems will generate health event rates around a putative point source and produce small area disease maps of any defined area. Whilst standard GIS software may be used to produce crude disease rates and maps, these can often be misleading (e.g. due to sparse data) and therefore more sophisticated statistical methods are needed to

create valid output. Thus, existing methods will be further developed to analyse and visualise the data in the most appropriate way for the level of data available.

3 Exploration of the use of the proposed system for Health Impact Assessments (January to December 2001).
Exposure maps produced by the proposed system may be used in conjunction with population data and existing exposure-response relationships to estimate the health impacts of defined environmental exposures.
Disease mapping may be used to monitor health outcomes related to sources of environmental exposure.

4 Social inequalities in health (January to July 2001)
Depending on the outcome of the feasibility study, structurally equivalent "deprivation" scores for small area analysis will be developed in the partner countries. The development of such scores will constitute an important aspect of the project.
Calibrating the modelling implications of different compositional forms of deprivation scores will be possible in several partner countries in order to quantify the sensitivity of the analysis to score composition.
A recommendation on a "best practice" in this context will result.

5 Develop protocols for case studies (September to December 2001)
Case studies will be designed to evaluate the implemented systems. The design should be available by the end of the project to enable the case studies to be performed during project year three.

6 Final report will be delivered 15 August 2002.

The report details the work of the each of the partners including the UK in separate chapters.

The last chapter includes EUROHEIS meetings and conferences held during the second year.

Appendix 1 consists of the documentation for the enhanced Rapid Inquiry Facility (RIF).

Note also that further details regarding the project can be obtained from the EUROHEIS web site (http://www.med.ic.ac.uk/divisions/60/euroheis/homepage.htm).

DENMARK

Arne Poulstrup

National Board of Health
Vejle
DENMARK

Background

The second phase of the EUROHEIS project aims to develop an integrated information system for assessment of relationships between environment and health at a geo-spatial-temporal level.

The integrated RIF-model developed in the UK has been the inspiration for the Danish decision-making process. However, it has been decided to go for a more comprehensive model in Denmark due to the fact that underlying health registers contain data on an individual level that can be linked with geo-coded addresses and datasets holding information on population, environment and socio-economy.

System

The information system is being developed using Arc View 8i and database DB2 8i. The aim is to create an information system – web based – with different levels of users with different levels of authorization. The front end will be linked to the database that will be updated periodically to ensure that the system does not loose its immediacy.

The basis for good decision-making, whether the concern is health surveillance or prevention are data of high quality, sufficient levels of security and meaningful linkage between environmental data and health and population data.
Furthermore it is necessary to define different levels of geographical units according to how sensitive data are for the purpose of disease mapping acknowledging that data are on an individual level.

Case studies

The development of the system is being tested with two case studies, one of which deals with the quality of drinking water and health outcome and the other dealing with contact allergy as a risk indicator. These two studies are based on very different types of environmental exposures – one affecting a relative small population within a relative small and well-defined area and the other affecting a large population in a large area. The two studies also include different types of health data – both central registered data (national databases) and data from local clinical databases.

The system will be utilised to define the population that has been exposed.
Development of the information system will include the provision of the appropriate environmental models to ensure correct estimations of exposure levels. Furthermore, development of the system will include standardisation of data and solutions to the import of data from different types of systems until a national environmental database has been developed. At the moment the environmental information lies on local and central databases with different set-ups, platforms and software facilities, different owners, different maintenance schedules, policies etc.

Depending on type of exposure the health outcome of the population will be assessed. Geo-referenced environmental data will be linked to geo-referenced health- and population data on an individual level identifying the exposed population. Development of the information system will include provision of the appropriate statistical models to ensure correct risk estimations and provide the user with appropriate tools for mapping diseases based on few and sparse data. Furthermore, the development of the system will include standardisation of data and solutions to import data from different types of medical databases.

Both for environmental- and health data the statistical basis will include methods for handling missing values and sparse data. Also the statistical basis will include the possibility of incorporating features as spatial correlation, cluster analysis and bias analyses. It will also be possible to describe the population demographically and link it to socio-economic variables.

The database will contain geo-referenced data on an individual level for health- and population data. The health data are derived from central public registers containing information on an individual level based on information from public and private hospitals.

It is important as well to develop methods for inclusion of data from local clinical databases, and from the primary healthcare sector, as many diseases related to environmental exposures are not diagnosed in or referred to the secondary health sector.

Population data are derived from the Central Population Register and then geo-coded. These geo-referenced data will also include historical demographic data, which makes it possible to estimate exposure levels not only on a spatial level but also on a temporal level which is important when focus is diseases with a longer period of latency.

The development of the system will at first be based on standard reports and disease mapping. The next step will be web based standard reports and the usage of disease mapping facilities and finally if a proper level of security can be obtained the system will be accessible for different user groups with a possibility of user defined reporting facilities. The user group will be limited to persons with the professional background and capability to handle and analyse data.

Development of the system is based on a close co-operation with a number of specialists within the fields of medicine and environment. It is seen of vital importance that the system is based on data of high quality and that methods and models are correctly incorporated in the system.

Development of the information system based on standard reports and disease mapping is following the stages mentioned below.

- Stage 1 (project description and contracts with suppliers of software and data)
- Stage 2 (purchase of soft- and hardware and training)
- Stage 3 (collation of data, maps and geo-coded addresses)
- Stage 4 (design and development of the GIS-model)
- Stage 5 (implementation, testing and calibration of the GIS-model)
- Stage 6 (documentation)

All tasks described in stages 1 to 3 are either completed or in progress.

Summary

Participation in the EUROHEIS project has been the start of a process that is thought to be an on-going process including development and refinement based on up to date methods and solutions. The future perspectives of the process will depend on how ambitious the information system is thought to be.

As can be noted, Denmark has decided not to replicate the UK RIF-model, but the RIF model and the EUROHEIS co-operation has been of great inspiration and of immense value in the decision-making and choice of the Danish model for fast environmental health surveillance, which hopefully will be in operation by the beginning of next year.

FINLAND

Esa Kokki

National Public Health Institute
Kuopio
FINLAND

Background

The aim of Phase II was in implementation of systems for point source investigations and disease and exposure mapping, modelled on RIF - the system being developed by the UK partner.

The other aim of Finnish partner was in the extension of statistical methods to solve the problems of fine resolution and sparse data in estimation of cancer risk around a point source.

Implementation

In Finland small area system called Small Area Statistics on Health (SMASH) (Kokki et al. 2001, Kokki et al. 2002A) is already in use. The Finnish system provides very similar analytic possibilities as the UK system and therefore there is little need to completely rebuild the Finnish system. However, implementation of the RIF has been started; licenses for the required software and the new 500 * 500 meters grid for geographical classification have been bought, and update of cancer up to year 2000 data is on order. The characteristics of the two systems will be compared.

Extension of statistical methods

A changepoint method is applied in estimation of cancer risk for zones around a point source (Kokki et al. 2002B). In this method, the number and the positions of changes in risk levels and estimates of the zone specific risks are the interesting parameters. The Poisson model for observed number of cases is exploited. The model is implemented by means of Gibbs sampling. The method was found useful; the positions of the changes in risk levels and corresponding risks can be found. The results about this and previously developed methods will be presented at the ISEA/ISEE 2002 conference (Kokki et al. 2002C).

Case study

In a case study in Phase three we will investigate the possible increase in cancer risk following exposure to chlorophenols and PCDD/F emanating from a heavily contaminated river. Total cancer, lung cancer, breast cancer, sarcoma, myeloma, and lymphoma will be studied. It is presumed that PCDD/Fs are mobilised from river sediments and accumulate in nearby residents either via contaminated drinking water or via the food chain. The exposure assessment will be based on the distance of the place of residence from the river and from the Gulf of Finland. The study area will be divided into nine sub-areas: by distance to the river (<1, 1-4, 5-19 km) and by distance to the sea (<20, 20-39, 40-59 km). The extremely high amounts of PCDD/F in the river are a potential public health problem.

References

Kokki E, Ranta J, Penttinen A, Pukkala E and Pekkanen J. (2001) Small area estimation of incidence of cancer around a point source of exposure with fine resolution data. Occupational and Environmental Medicine 58, 315-320.

Kokki E, Pukkala E, Verkasalo P, Pekkanen J. (2002A) Small Area Statistics on Health (SMASH) - A System for Rapid Investigations of Cancer in Finland. In Briggs D, Forer P, Jarup L, Stern R (eds.): Geographical Information Systems (GIS) for Emergency Preparedness and Health Risk Reduction, Kluwer Academic Publishers, Dordrecht. (in press).

Kokki E, Penttinen A. (2002B) Estimation of risk of disease around a point source with changepoint method. (Manuscript).

Kokki E, Penttinen A, Pukkala E, Verkasalo PK, Pekkanen J. (2002C) Small area statistics on health (SMASH) - a system for rapid investigations of cancer in Finland. ISEA/ISEE 2002: Linking exposures and health: innovations and interactions. Vancouver, Canada, 11.-15.8.2002. (Conference abstract).

IRELAND

Dr. Alan Kelly
Mr. Conor Teljeur

Small Area Health Research Unit
Trinity College Dublin
IRELAND

Project aim

To develop and test appropriate and sensitive measures of deprivation at a suitable spatial level (small-area level for preference) in partner countries. In order to achieve this, we would need to determine the relationship between various socioeconomic measures and health outcome data at the small-area level from selected partner countries. It was expected that the data described would presently be held by the partner or may be readily obtained from public sources in the partner's country.

Deprivation and Health – brief relevant background

Deprivation is a concept that has taken a variety of forms and has had many different meanings that have evolved over time. It is generally recognised as a composite concept, in that there is no single variable that can be said to measure it but rather a number of variables must be combined in some way. While poverty, as measured by household income, is usually recognised as an important component of deprivation it is only one of many variables affecting quality of life. Moreover the measurement of poverty remains controversial, governments tend to dislike the political implications and the poor dislike the stigma of being labelled poor.

Deprivation has been defined as a state of "observable and demonstrable disadvantage relative to the local community to which an individual belongs". The idea has come to be applied to conditions (i.e. physical and social circumstances) rather than resources or income and can therefore be distinguished from the concept of poverty, though the two are closely related. This conceptualisation can explain why people can experience deprivation but do not necessarily live in poverty.

The measurement of deprivation has been pursued energetically in the UK since the early 1980s and a number of deprivation indices have been put forward. Some of these indices have been defined specifically in relation to health while others, designed in a different context, have been appropriated into the health field. Early attempts at deriving suitable deprivation indices exhibited many methodological differences and the number of indicators included were large, typically ranging from 8 - 12 items. One recent approach to measuring deprivation attempts to locate areas (and the populations in them) on a dimension which reflects the access people have to goods, services, resources, amenities and physical environment, which are customary, or at least widely aspired to in society. This approach has lead to the compilation of "material" deprivation indices that have gained considerable credence in the UK, not least for their sound conceptual basis. Two material deprivation indices have been developed in the UK, one for the North of England (the so-called 'Townsend Index) and the other for Scotland ("Carstairs Index").

Since 1997, Ireland has also had a national deprivation index developed by SAHRU along the lines of the Townsend and Carstairs indexes in the UK.

In the present phase of the EUROHEIS 3-year project, we proposed to 1) determine the availability and scope of socioeconomic and health data within the consortium, and 2) explore the country specific relationships between measures of socioeconomic status and health outcomes at an aggregated level.

The Meta Data exercise

We undertook an extensive review of meta data protocols for the exchange of meta data between partners in relation to the health and socio-economic databases available to the partners. The standard chosen – in view of its better support for spatially-referenced data – is that employed in the UK (The Meta Data Exchange Protocol).

An Excel file containing VBA macros was developed based on the UK Meta Data Exchange Protocol that would be used to create a database of MetaData held or available to each partner. The Excel file prompted the user for the specifications of each dataset that can be in one of three categories ('health', 'socioeconomic' or 'other'). The user can fill in details such as data origin, time period of data, spatial level and various additional variables associated with each dataset. When all relevant datasets have been entered by a partner the Excel file would be returned to us for assessment. The file was compatible with both PC (Windows 95, 98 and 2000) and Mac.

Distribution

A copy of the 'MetaData.xls' file (available on request from Mr. Conor Teljeur at teljeurc@tcd.ie) was sent out to each partner country in August 2001 along with a descriptive text on how to use the file. An initial response from Sweden related to a problem with compatibility that was eventually resolved. By early 2002 full responses had been received from Denmark, Italy, Finland and Spain. Sweden delayed (due to Christina Reuterwall's move to Stockholm) and will respond in due course. A response was not requested from the UK partner, as details of data held at SAHSU were available.

Summary of Metadata Report

All data referred to below are available at small area (or below) level with the exception of Italy which is at municipality level.

Denmark

Has socio-economic data covering education, occupation and income which may be sufficient to compile a deprivation index
Has all cause mortality and cause specific (cancer) morbidity data.

Finland

Has information on health data (cancer incidence)
Has socioeconomic index

Italy
Has an index of socio-economic deprivation comprised of education, housing quality, unemployment and family structure variables
Has cause specific mortality data at a regional level

Spain
Has numerous socio-economic variables that can be used to develop a deprivation index (including unemployment, occupation, income, illiteracy and vehicle ownership).
Has all cause mortality and cause specific morbidity data

Ireland
Has a national deprivation index
Has cause specific mortality data

UK
Has a number of national deprivation indexes
Has cause specific mortality data

As Italy already has a deprivation index, the principle task with regard to their data will be a straight analysis of deprivation and health outcomes. With regard to Denmark and Spain, there are possibilities for the development of deprivation indices that will be followed by comparisons with health outcomes. The Finnish data has a derived socioeconomic variable already. We have no response as yet from Sweden, but we understand that relevant social and economic data are available.

Analysis

Data were requested from Spain and Finland (in addition to Irish data available to SAHRU) with a view to exploring country specific associations between socioeconomic variables and health outcomes. A priori, it is assumed that such general relationships exist (as discussed in our end of Phase I report) in all partner countries; it is not assumed that the nature and strength of such relationships would be the same in these countries.

Extensive and detailed analysis of country data have (and continue) to be undertaken to understand the evidence of any socioeconomic relationship found with health outcomes. The variables available within partner datasets differ significantly (not unexpectedly), as does the outcome variables. In view of this diversity, analyses conducted to-date has therefore had to focus on extracting whatever within-country patterns the available data suggest before common cross-country themes can be sought. Additional data will be sought from other partners – as needed – to further explore these relationships and to attempt to extract some general conclusions appropriate to the various partner counties with such a heterogeneous economic and heath experience. In Phase III, preliminary findings will be discussed with the partners and their input sought in attempting to understand and explain findings. This would be preparatory to assisting partners in designing a deprivation index that meets their particular needs.

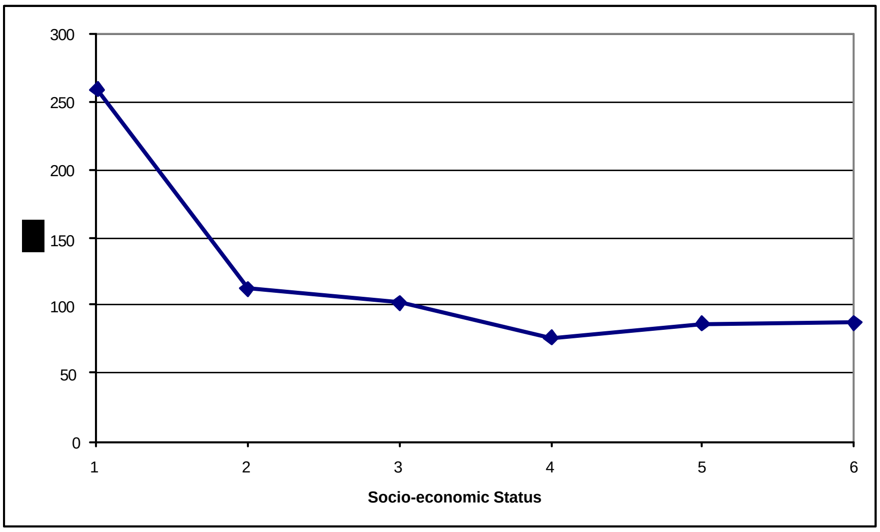
Review of country data

In the following pages, we have put together some straightforward analyses of the relationships between socioeconomic factors and health outcomes based on the data supplied by our Spanish and Finnish colleagues, as well as some results from a similar Irish analysis. The more detailed spatial analysis in on-going.

Finland

Finland has provided datasets relating to regional population and regional cancer incidence. Information in both datasets is coded to x- and y spatial grid coordinates. Both datasets include a socio-economic status variable and municipality code. For this overview, data have been aggregated for the whole area to calculate age-sex standardised incidence (SIR) rates for all cancers by socio-economic status.

Cancer SIRs by socio-economic status in Finland.



There is a clear and marked variation in cancer morbidity across different groups of socio-economic status, with significantly higher rates in the lower (SES 1) socioeconomic class compared to the remaining classes. Standardised rates in classes 2 through 6 do not differ significantly from each other.

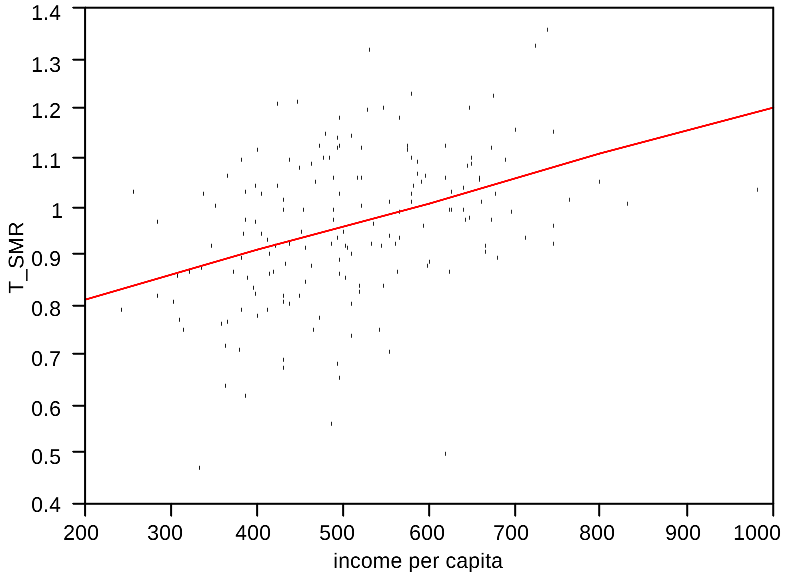
Spain

Spain provided vary detailed data for two different regions: Granada and Valencia. The data for Granada contains mortality (all cause and breast cancer) and a range of socioeconomic variables (including so-called 'aged rate', per capita income, illiteracy rates, unemployment rates, activity rate and a rural-urban indicator). All data are at the municipality postcode level.

The Granada postcoded mortality figures were adjusted using empirical Bayes methods before fitting bivariate models . The results are somewhat counter-intuitive and await discussions with our Spanish partner before a more detailed analysis and interpretation can proceed.

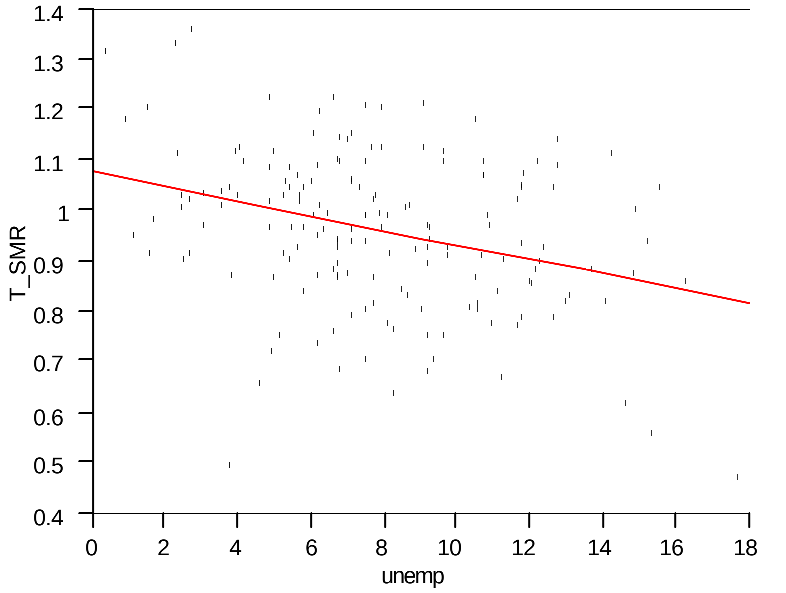
All cause mortality: Granada

All Cause SMR by per capita income



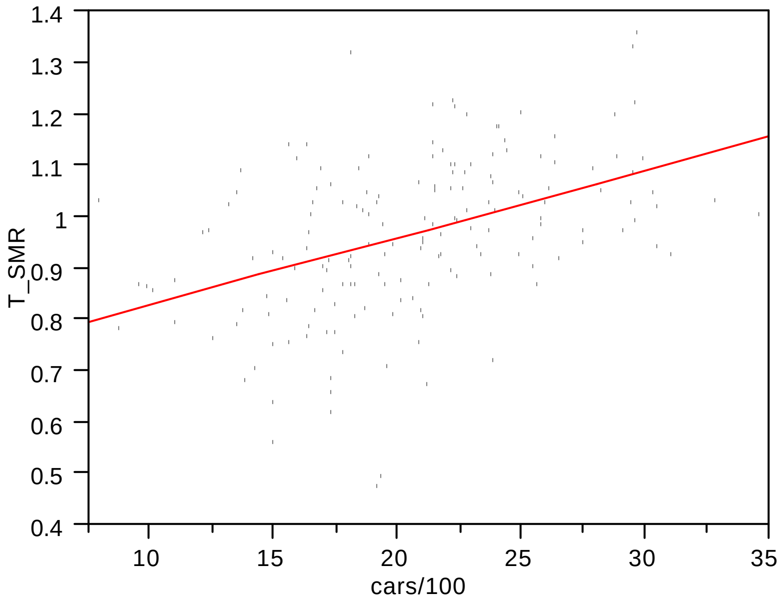
Regression line:
All cause SMR = 0.7197068 + 0.0004799 income per capita
P < 0.0001

All Cause SMR by unemployment rate



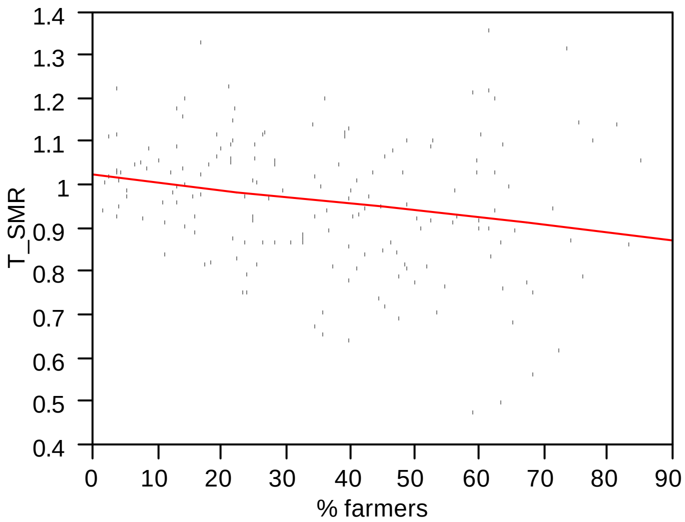
Regression line:
All cause SMR = 1.0756725 - 0.0143479 unemployment rate
P < 0.0001

All Cause SMR by cars per100 persons



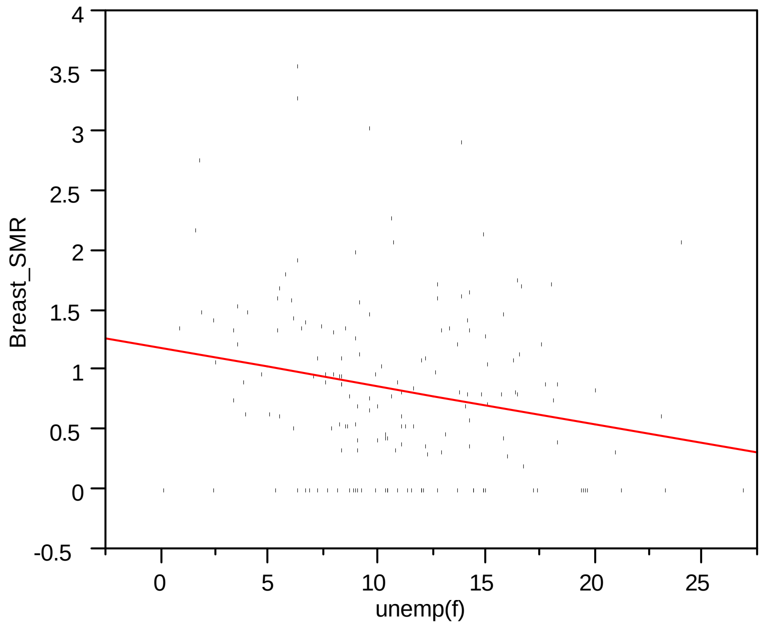
Regression line:
All cause SMR = 0.6968914 + 0.013023 cars/100 persons
P < 0.0001

All Cause SMR by percentage farmers



Regression line:
All cause SMR = 1.021394 - 0.0016288 percentage farmers
P < 0.0036

Breast cancer SMR by female unemployment rate

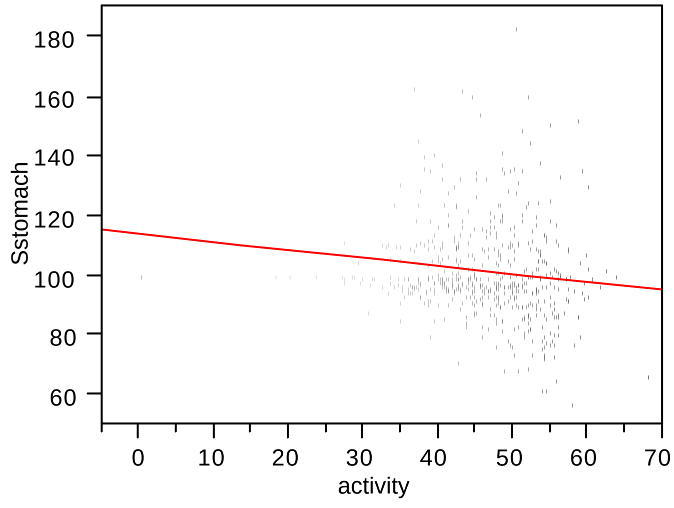


Regression line:
Breast_SMR = 1.1645104 - 0.0308666 female unemployment rate
P < 0.0062

Valencia

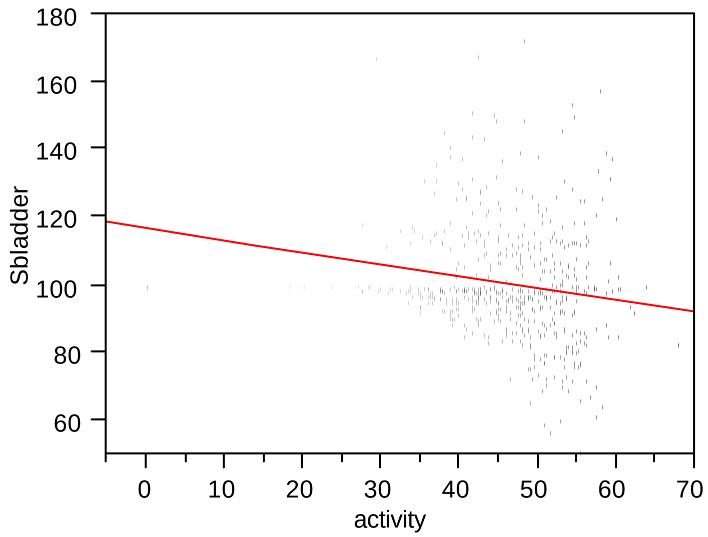
The Valencia data also includes SMRs (for leukaemia, cancer of the stomach, bladder and prostate, cerebrovascular and ischaemic heart disease) and a number of socio-economic variables. The proportion of workers in different occupational groups is given along with illiteracy and activity rates. Several water quality measures have also been supplied.

SMR (stomach cancer) by activity rate



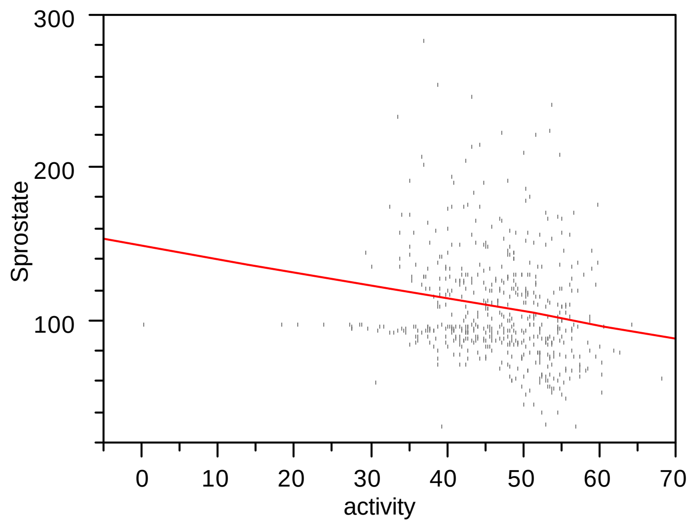
Regression line:
SMR(stomach cancer) = 113.88581 - 0.2723221 activity rate
P < 0.0022

SMR (bladder cancer) by activity rate



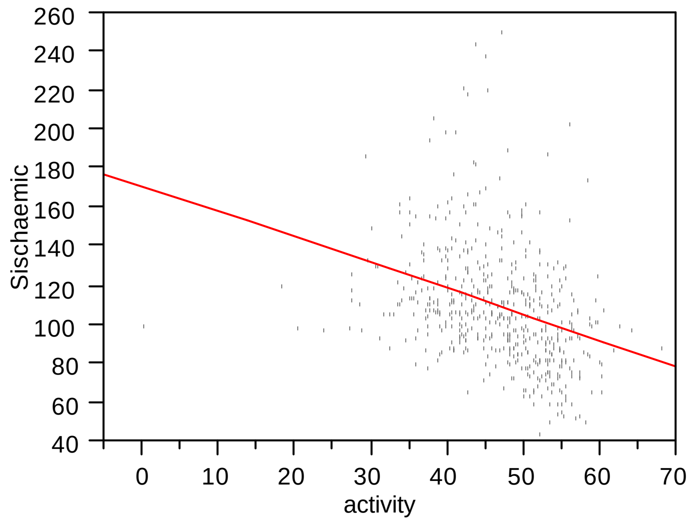
Regression line:
SMR (bladder) = 116.53224 - 0.3485825 activity rate
P < 0.0002

SMR (prostate cancer) by activity rate



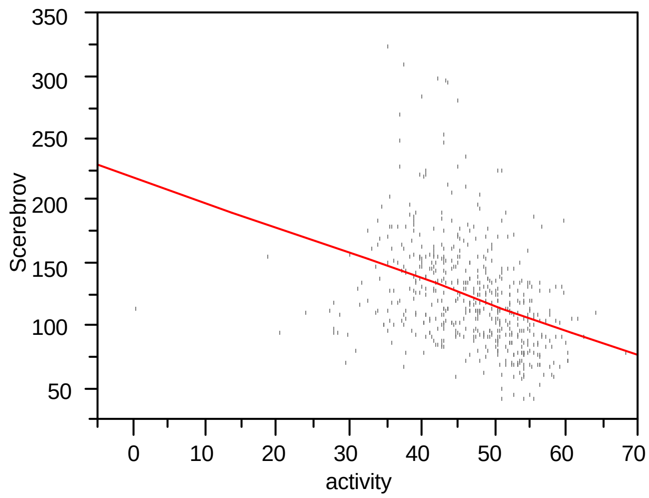
Regression Line:
SMR (prostate) = 148.74346 - 0.8567715 activity rate
P < 0.0001

SMR (ischaemic heart disease) by activity rate



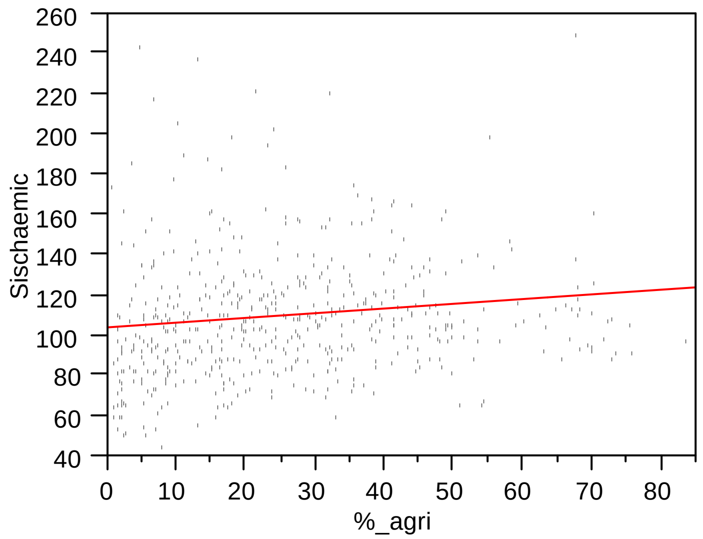
Regression line:
SMR (ischaemic) = 170.83414 - 1.3292883 activity rate
P < 0.0001

SMR (cerebrovascular disease) by activity rate



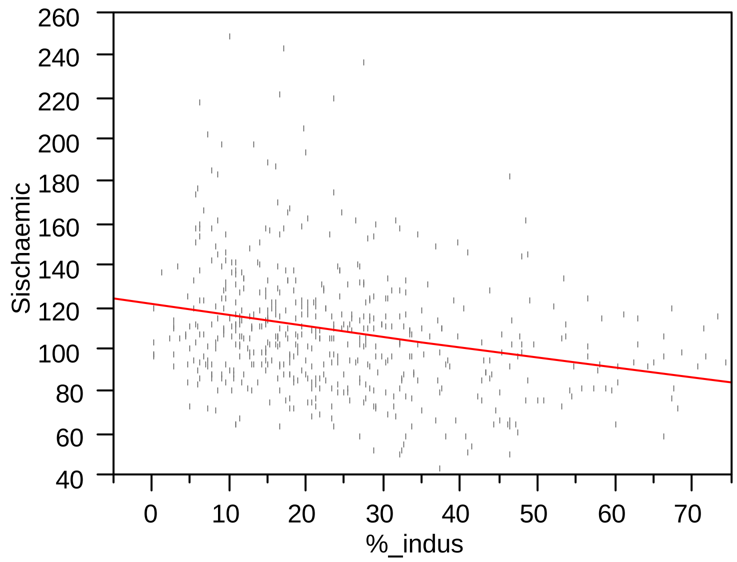
Regression line:
SMR (cerebrovascular) = 218.44412 - 2.016264 activity rate
P < 0.0001

SMR (ischaemic heart disease) by % agricultural workers



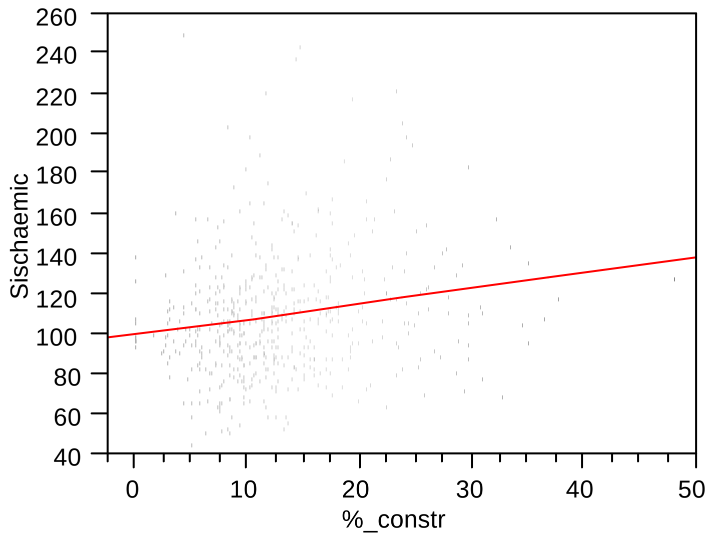
Regression line:
SMR (ischaemic) = 103.29936 + 0.236799 percent agriculture workers
P < 0.0011

SMR (ischaemic heart disease) by % industrial workers



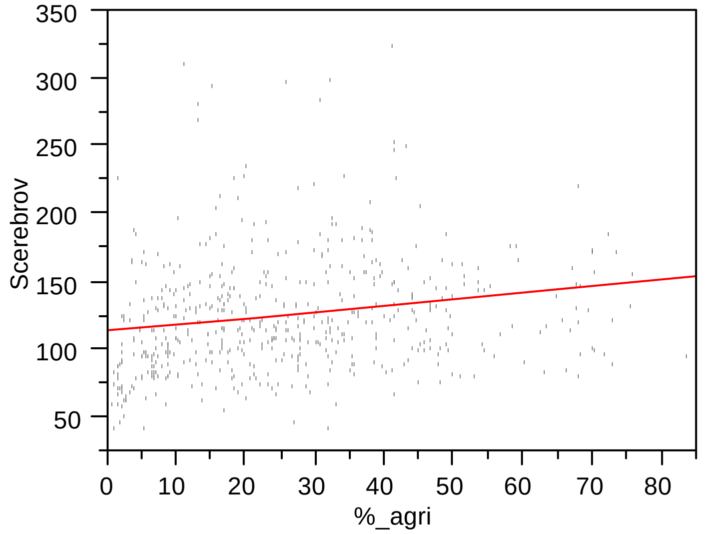
Regression line:
SMR (ischaemic) = 121.30614 - 0.4967681 percent industrial workers
P < 0.0001

SMR (ischaemic heart disease) by % construction workers



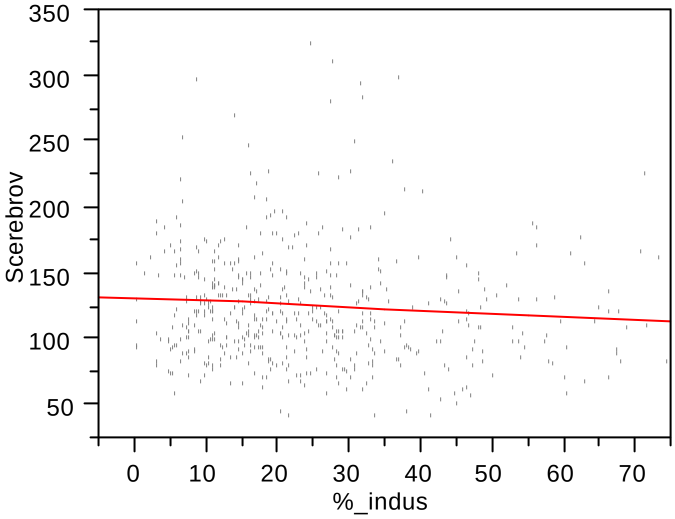
Regression line:
SMR (ischaemic) = 99.822013 + 0.7499005 percent construction workers
P < 0.0001

SMR (cerebrovascular disease) by % agricultural workers



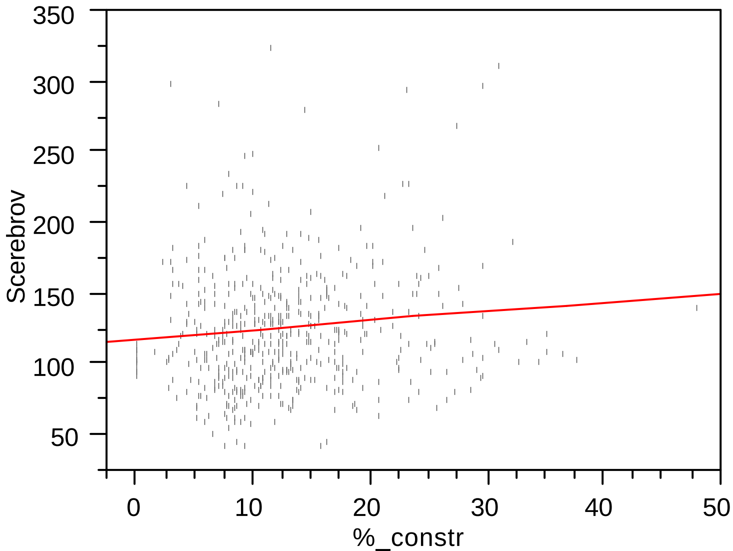
Regression line:
SMR (cerebrovascular) = 113.19352 + 0.4696557 percent agricultural workers
P < 0.0001

SMR (cerebrovascular disease) by % industrial workers



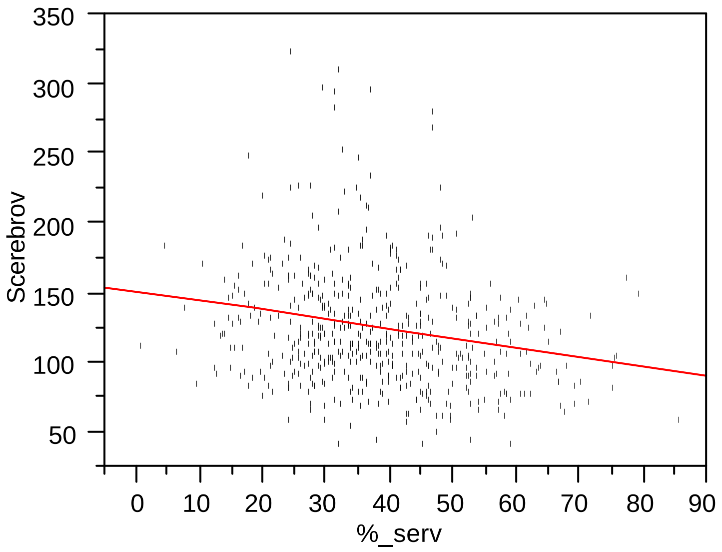
Regression line:
SMR (cerebrovascular) = 130.69238 - 0.2311389 percent industrial workers
P < 0.0499

SMR (cerebrovascular disease) by % construction workers



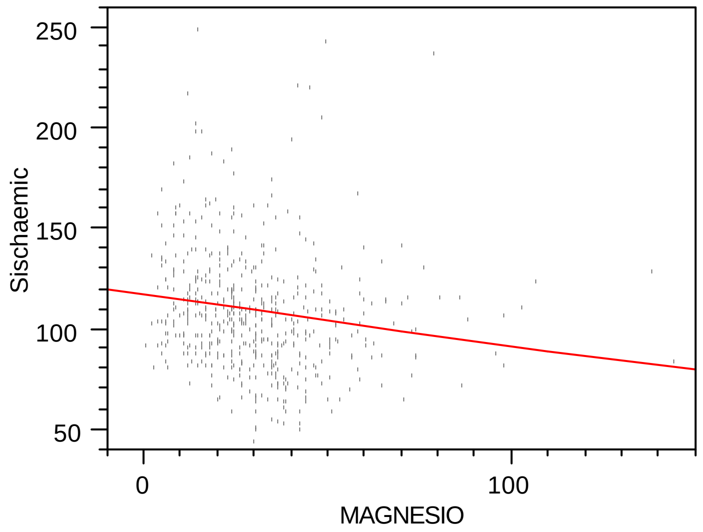
Regression line:
SMR (cerebrovascular) = 116.57773 + 0.6745129 percent construction workers
P < 0.0095

SMR (cerebrovascular disease) by % services workers



Regression line:
SMR (cerebrovascular) = 150.37007 - 0.667814 percent service workers
P < 0.0001

SMR (ischaemic heart disease) by level of magnesium



Regression line:
SMR (ischaemic) = 116.8255 - 0.2521296 level of magnesium
P < 0.0003

Principal Components analysis

As a preliminary indication of the potential of deriving a deprivation from the data for Granada and for Valencia, we have undertaken a principal components analysis of the potential set of explanatory variables supplied for each region separately.
(Correlations are given at the end of this section)

Results for Granada

Principal Components: (based on correlation matrix)

EigenValue	Percent	Cum Percent
3.5938	35.938	35.938
1.3760	13.760	49.698
1.2730	12.730	62.428
1.1141	11.141	73.569
0.7835	7.835	81.404
0.5940	5.940	87.343
0.4678	4.678	92.022
0.3471	3.471	95.492
0.2783	2.783	98.275
0.1725	1.725	100.000

The first 3 principal components:

Eigenvectors	1st	2nd	3rd
Aged rate	-0.33757	-0.04338	-0.43156
income per capita	0.42543	0.05810	-0.28735
Illiteracy rate	-0.15538	-0.55907	0.44964
RuralUrban	-0.30009	-0.22399	-0.44796
%growing surface	0.09879	-0.15041	0.21437
%greenhouse	0.01107	0.59190	0.17675
%unemployment	-0.27820	0.23011	0.48389
activity level	0.42688	0.16344	0.02768
cars/100 pop.	0.45304	-0.11470	-0.05910
% farmers	-0.34075	0.40770	-0.11284

Several variables score about the same on the 1st PC: aged rate, income per capita, activity level, cars/100 population and % Farmers. Illiteracy rate and % greenhouse score high on the 2nd PC. Aged rate, Illiteracy rate, Rural/Urban and %Unemployment score equally on the 3rd PC.

Results for Valencia

Principal Components: on Correlations

EigenValue	Percent	Cum Percent
2.0271	33.786	33.786
1.3685	22.808	56.593
1.0179	16.965	73.558
0.9603	16.004	89.562
0.6263	10.438	100.000
0.0000	0.000	100.000

The first 3 principal components:

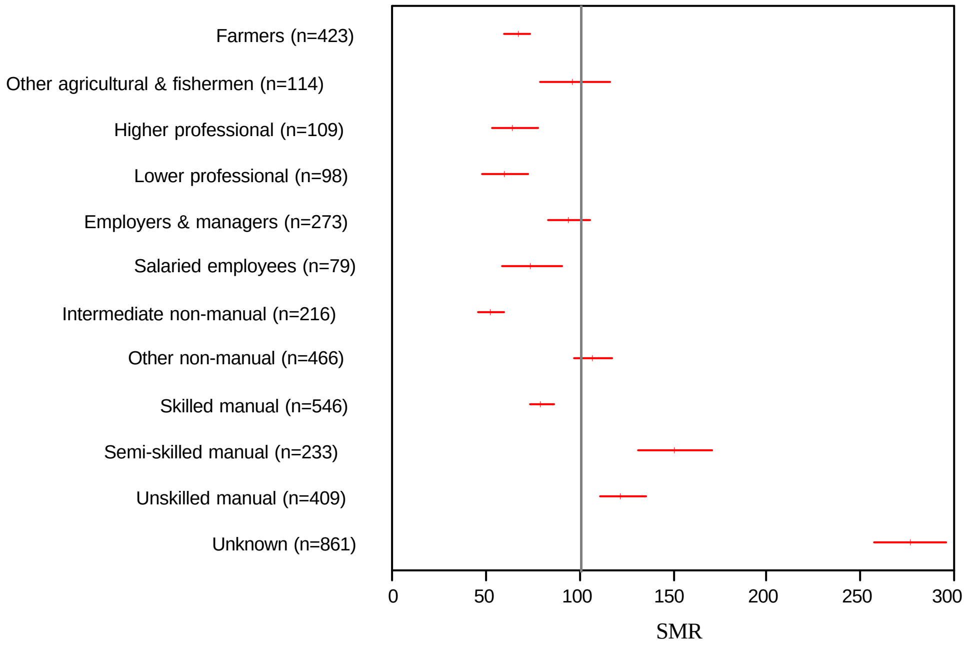
Eigenvectors	1st	2nd	3rd
Activity level	0.47703	-0.05137	0.36469
illiteracy	-0.18876	-0.04868	0.21364
%_agriculture	-0.62850	-0.14408	0.30947
%_industry	0.41751	-0.58985	-0.37511
%_construction	-0.10703	0.53936	-0.66983
%_services	0.39499	0.57914	0.36906

Activity level, % agriculture (significantly higher than:) % industry and % services score high on the 1st PC. % Industry, % construction and % services score equally high on the 2nd PC. The 3rd PC is dominated by % construction. %

Ireland

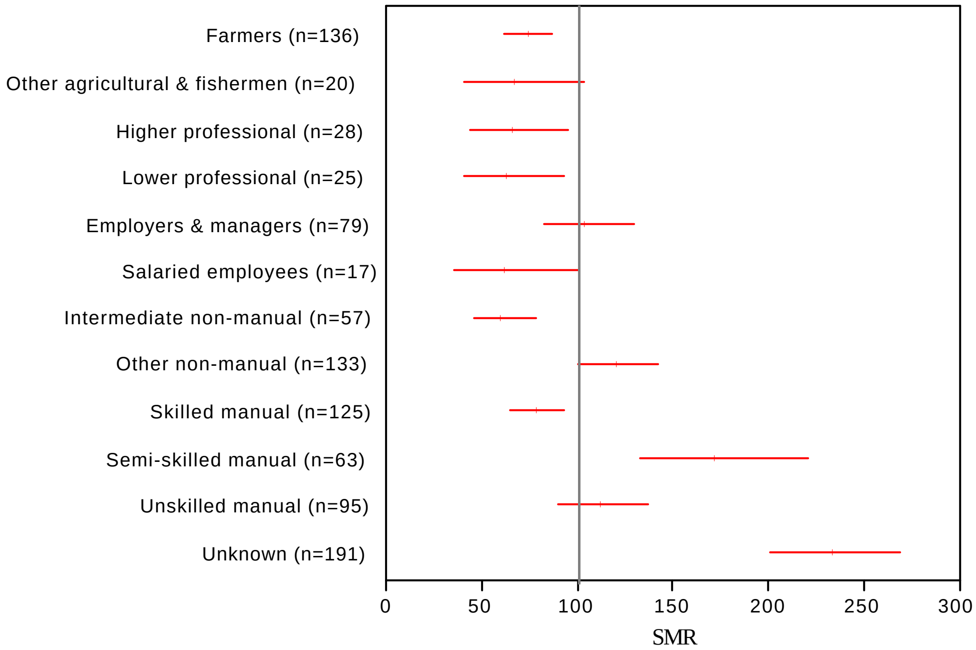
The following plots are extracted from a recent report 1 produced by colleagues in the Department of Community Health & General Practice in Trinity College Dublin. The analyses show the relationship between SMRs and socioeconomic group for All Cause mortality, Ischaemic Hearth Disease and Neoplasms.

Standardised Mortality Ratios (SMRs) and 95% confidence intervals
for all causes (males 15-64 years) by socio-economic group,
Ireland 1996

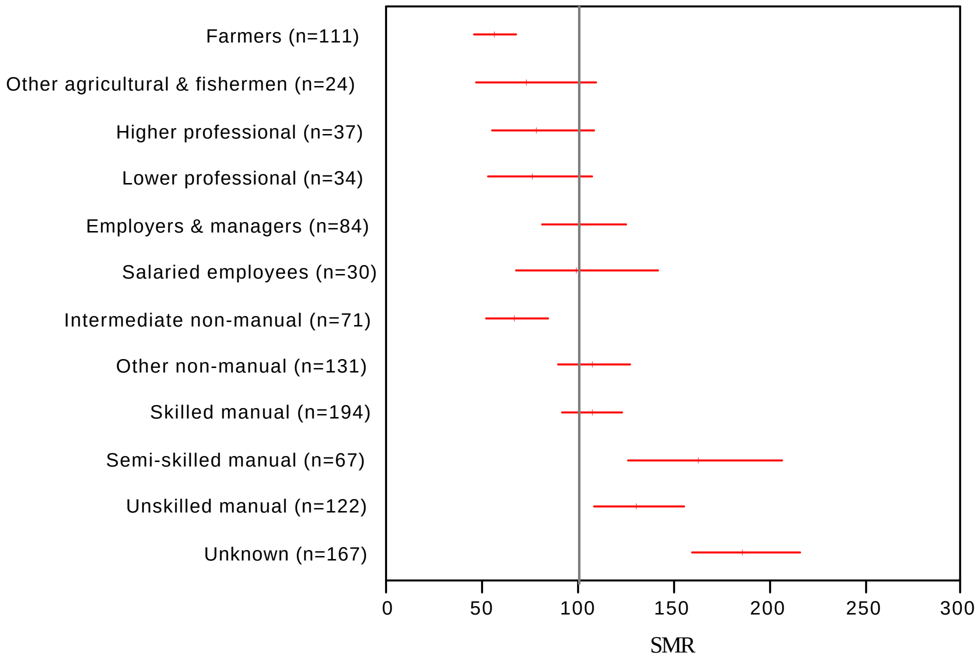


[1] Inequalities in Health in Ireland – Hard Facts, Barry, J., Sinclair, H., Kelly, A. et al.

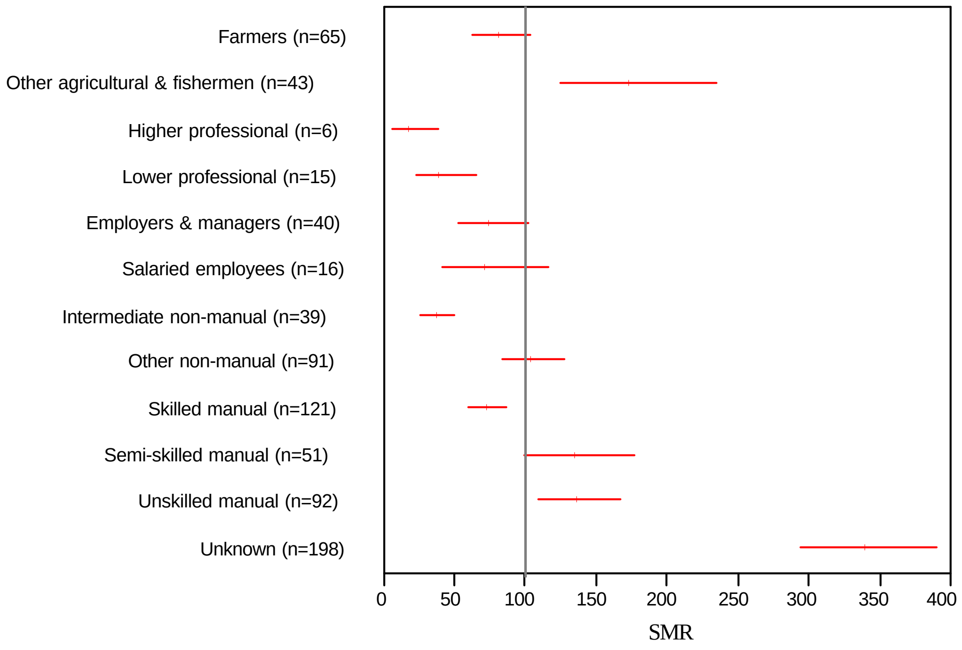
Standardised Mortality Ratios (SMRs) and 95% confidence intervals
for ischaemic heart disease (males 15-64 years) by socio-economic group,
Ireland 1996



Standardised Mortality Ratios (SMRs) and 95% confidence intervals
for neoplasms (males 15-64 years) by socio-economic group,
Ireland 1996



Standardised Mortality Ratios (SMRs) and 95% confidence intervals for injuries and poisonings (males 15-64 years) by socio-economic group, Ireland 1996



Conclusions

Only preliminary findings are presented in this report as data were provided to this partner rather late in the Phase II project period. Detailed analyses are progressing now, including spatial modelling and additional country data are being sought from remaining partners in the coming months.

What is immediately clear from the results to hand, will be the need to tailor any index of area-level deprivation to the individual country or region within a country. We are able to say that, by-and-large, trends in standardised mortality are broadly in accordance with general economic status trends in Spain, Finland and Ireland (although note the exception in Granada), however, this generalisation was fully anticipated, although not previously quantified, as here. The principal components analysis for Granada and Valencia offers promise for developing a weighted sub-set of available area-level descriptors in a form suited to a deprivation index – but again, it is likely that we would consider two indexes would be required rather than a single index covering both areas.

Much remains to be done. The primary task requires a detailed exploration of the presently available data and new data from our remaining partners with a view to identifying - for each partner - a sub-set of socioeconomic variables that best discriminates between high and low mortality or morbidity rates on a small area basis in order to provide for the possibility of adjusting for potential confounding in the analysis of environmental and health data.

Correlations: Granada data

	Aged rate	income per capita	Illiteracy rate	RuralUrban	%growing surf	%greenhouse	unemployment	activity level	cars/100	% farmers
Aged rate	1.0000	-0.2616	0.0645	0.4679	-0.1643	-0.0871	0.1455	-0.5017	-0.4542	0.3998
income per capita	-0.2616	1.0000	-0.3513	-0.3079	0.0663	0.0329	-0.4381	0.5954	0.7481	-0.4294
Illiteracy rate	0.0645	-0.3513	1.0000	0.1259	0.0521	-0.1418	0.1953	-0.3044	-0.1828	-0.2041
RuralUrban	0.4679	-0.3079	0.1259	1.0000	-0.1367	-0.1034	0.0005	-0.4721	-0.3775	0.2423
%growing surface	-0.1643	0.0663	0.0521	-0.1367	1.0000	-0.0451	-0.0918	-0.0316	0.3211	0.0510
%greenhouse	-0.0871	0.0329	-0.1418	-0.1034	-0.0451	1.0000	0.1211	0.0930	-0.0430	0.1463
unemp	0.1455	-0.4381	0.1953	0.0005	-0.0918	0.1211	1.0000	-0.2993	-0.4492	0.3422
activity level	-0.5017	0.5954	-0.3044	-0.4721	-0.0316	0.0930	-0.2993	1.0000	0.5478	-0.4880
cars/100	-0.4542	0.7481	-0.1828	-0.3775	0.3211	-0.0430	-0.4492	0.5478	1.0000	-0.5041
% farmers	0.3998	-0.4294	-0.2041	0.2423	0.0510	0.1463	0.3422	-0.4880	-0.5041	1.0000

Correlations: Valencia data

	activity	illiteracy	%_agri	%_indus	%_constr	%_serv
activity	1.0000	-0.1104	-0.3665	0.2454	-0.1899	0.2937
illiteracy	-0.1104	1.0000	0.1280	-0.0743	-0.0077	-0.0774
%_agri	-0.3665	0.1280	1.0000	-0.5782	-0.1521	-0.5606
%_indus	0.2454	-0.0743	-0.5782	1.0000	-0.2807	-0.2422
%_constr	-0.1899	-0.0077	-0.1521	-0.2807	1.0000	0.0020
%_serv	0.2937	-0.0774	-0.5606	-0.2422	0.0020	1.0000

ITALY

Marco Martuzzi

WHO European Centre for Environment & Health
Rome
ITALY

Background

Health impact assessment (HIA) is of growing interest in the field of public health in Europe. Scientific knowledge on the adverse effects of several environmental factors on the population is in some cases substantial, but often regulatory policies fail to reflect such knowledge adequately. In addition the Amsterdam Treaty of the European Union (Article 152) states that definition and implementation of all new Community policies should ensure a high level of human health protection, which can be achieved through HIA.

A key element to support the formulation of public health policies based on the available evidence is to develop rigorous methods to translate experimental, toxicological and epidemiological information into accurate estimation of the overall impact on the health of the population. So far, most of these exercises have been based on relatively simplistic methods, where the health impact is measured by direct derivation of attributable risks, based on available dose-response estimates and exposure profile of the population (often an average measure), and not using statistical models. Uncertainties in the estimates of the impact are normally evaluated on the basis of the confidence intervals attached to the dose-response function, and other sources of uncertainty that can be of great importance are ignored, or acknowledged only qualitatively. In addition, concerns of "double counting" of events in multiple health endpoints assessments, often suggest limiting studies to a partial set of adverse health consequences. As a result, impact assessments are often said to be "conservative", and little or no attempt is made to verify that the necessary assumptions are met and/or to carry out quantitative evaluation of the true systematic and random error involved.
Using data on urban air pollution and health, statistical methods based on a Bayesian approach have been explored. Assessment of health risks related to air pollution levels for each city has been calculated.

This work suggests that model-based methods for health impact assessment can be valuable to derive more accurate estimates of health impact, thus enhancing the scientific basis for decision making.

Methods

To estimate urban population exposure from the existing monitoring data, relevant monitoring stations for the eight major Italian cities have been selected and daily pollutants data from 1998 and 1999 have been considered. An average value from the two years empirical distributions has been calculated and used as proxy of the current population exposure. When available, values from stations measuring PM10 have been used; otherwise, estimates have been made using TSP data, applying correction coefficients derived from literature or directly calculated from available Italian data. To calculate the attributable proportion of health effects from air pollution, a relative risk value has been obtained by pooling the estimates of the available studies through a meta-analytic approach. Conservative dose-response coefficients have been selected to estimate the attributable number of cases.

Conventional models

The model normally applied to calculate the attributable proportion (A) of health effects of a given exposure to a risk factor (air pollution in the example discussed here) is the following one:

$$\textbf{(1)} \qquad A = (RR - 1) / RR$$

Where RR is the Relative Risk obtained from the available literature.
The central estimate of the necessary relative risks and 95% confidence bounds were derived using a meta-analytic approach.

To calculate the number of cases attributable to air pollution (E), the following formulation is used:

$$\textbf{(2)} \qquad E = A * B * C * P$$

Where **B** = population baseline rate of the given health effect
C = relevant change in exposure
P = relevant exposed population for health effect

The population baseline rate is the proportion of the exposed population that would experience the health outcome assuming a baseline level (or no effects level) of air pollution.

This can be calculated as:

$$\textbf{(3)} \qquad B = Bo / [1 + (RR-1)C]$$

Where **Bo** = observed rate of the health effect under current exposure
B = baseline rate of the health effect under baseline exposure
C = relevant change in exposure
RR = relative risk estimate.

Bo and **P** are obtained from available health statistics and from census data.

The Bayesian model

A Bayesian probability model, in which all the quantities are stochastic variables, was defined and applied to the data.

The model consists of a full joint probability distribution of all the unobserved (i.e. parameters and missing values) and observed quantities (data). This distribution is conditioned by data and posterior distributions for parameters and unobserved quantities are obtained. From each of these h conditional posterior distributions, alternatively for each unknown parameter and keeping fixed the remaining ones, a series of random values is generated through a large number (n) of simulations. It has been demonstrated that, under general conditions, every set of simulations can be considered as a sample extracted from the joint distribution of probability. The estimate for the unknown parameter is derived from these values through descriptive statistics (means, medians, percentiles).

The algorithm used to estimate the unobserved quantities belongs to the family of Monte Carlo simulation iterative methods and it is a procedure of numerical integration known as Gibbs sampling.

Through this Monte Carlo simulation method it is possible to calculate attributable risks and number of cases with confidence intervals for every single city and for a joint variable "sum" of all the cities. By using this Bayesian approach the width of the intervals also depends on the variability of the eight cities empirical distributions of pollutants.

Every city contributes to the variability of the full model with its own variability. To get the final values for every single city and for the total of the cities a sufficient number of simulations (10 000, without considering the first 1 000) is done through the Monte Carlo method described above.

Details of the models, and procedures followed for model fitting are given in Appendix 2.

Results

Several models were used to explore some of the available options, as follows.

Model 1: Pr 24 new data - variability of Relative Risk; use of arithmetical mean;

Model 1(a): Same variability of Model 1; confidence interval calculated in the correct way;

Model 2: WinBugs, first version - same variability of Model 1; variability of pollutants distribution not included; use of arithmetical mean;

Model 3: WinBugs, second version – same variability of Models 1 and 2 plus variability of the pollutants distribution; use of arithmetical mean;

Models 4: WinBugs, last version - same variability of Model 3; use of geometrical mean; first 1 000 iterations thrown away.

The results obtained using the above-mentioned different approaches are shown in Figure 1. The tables illustrate the number of attributable cases estimated by the different models, using three different reference levels for PM10 concentrations: 20, 30 and 40 µg/m3. It can be seen that central estimates are not substantially affected by the choice of the model. For example, using 30 as a reference value (first table), all models estimate that around 5,000 extra deaths are attributable to PM10 in the eight Italian cities each year. While this kind of information is valuable to describe the health benefits associated with abatement measures (and can therefore help formulate evidence-based policies), it is also crucial to be able to evaluate the degree of uncertainty that surrounds the estimates. The tables show the 95% confidence intervals attached to the estimates. Unlike the central estimates themselves, confidence limits are heavily affected by the choice of the model: in particular, models that allow for more sources of variability produce wider confidence intervals.

Figure 1. Comparison of results among different Model.

Baseline 20 - sum of the 8 cities	mean	IC 95% (LL)	IC 95% (UL)	2.5 perc	median	97.5 perc	CI width	N° of Simul.	Notes	Type of mean used
1 - No Bugs: var of RR	5143	1900	7990	-	-	-	6090	0	Variability of RR	Arithmetical
1a – No Bugs: var of RR (correct)	5143	3890	6396	-	-	-	2505	0	Variability of RR	Arithmetical
2 - Bugs: var of RR	5105	-	-	3768	5120	6384	2616	10000	Variability of RR	Arithmetical
3 - Bugs: var of RR and C	4812	-	-	1110	4787	8642	7530	10000	Variability of RR and C	Arithmetical
4 - Bugs: var of RR and C	5018	-	-	1909	4752	9708	7799	10000*	Variability of RR and C	Geometrical

Baseline 30 - sum of the 8 cities	mean	IC 95% (LL)	IC 95% (UL)	2.5 perc	median	97.5 perc	CI width	N° of Simul.	Notes	Type of mean used
1 - No Bugs: var of RR	3496	1272	5509	-	-	-	4237	0	Variability of RR	Arithmetical
1a – No Bugs: var of RR (correct)	3496	2598	4394	-	-	-	1797	0	Variability of RR	Arithmetical
2 - Bugs: var of RR	3479	-	-	2540	3488	4391	1851	10000	Variability of RR	Arithmetical
3 - Bugs: var of RR and C	3158	-	-	-745	3151	7026	7771	10000	Variability of RR and C	Arithmetical
4 - Bugs: var of RR and C	3376	-	-	2046	3107	8198	6152	10000*	Variability of RR and C	Geometrical

Baseline 40 - sum of the 8 cities	mean	IC 95% (LL)	IC 95% (UL)	2.5 perc	median	97.5 perc	CI width	N° of Simul.	Notes	Type of mean used
1 - No Bugs: var of RR	1765	632	2823	-	-	-	2191	0	Variability of RR	Arithmetical
1a – No Bugs: var of RR (correct)	1765	1269	2262	-	-	-	992	0	Variability of RR	Arithmetical
2 - Bugs: var of RR	1764	-	-	1248	1767	2257	1009	10000	Variability of RR	Arithmetical
3 - Bugs: var of RR and C	2051	-	-	-2779	1453	5340	8119	10000	Variability of RR and C	Arithmetical
4 - Bugs: var of RR and C	1642	-	-	-1800	1387	6552	8332	10000*	Variability of RR and C	Geometrical

*In the last model developed with WinBugs, 10 000 simulations have been made but the first 1 000 have been thrown away.

Discussion

Exploratory work carried out during Phase II of EUROHEIS indicates that it is feasible to apply methods that allow estimating attributable risks with appropriate treatment of several sources of systematic and random error. The methods have been explored using real data from a participating country. The methodology has been identified and tested on these data. Results suggest that it is possible to develop a tool that can be used in conjunction with data on environment and health in order to: further assist health professionals involved with HIA; facilitate the evaluation of the concrete public health implications; allow effective communication with the general public and decision makers alike; and ultimately assist in the development of appropriate policies.

The methods presented here should be further developed through applications in different case studies, to be identified within the EUROHEIS collaboration. In addition, areas of collaboration with the APHEIS project have been explored. APHEIS has, among its objectives, the aim of establish mechanisms for the monitoring of air pollution-related effects across Europe, including health impact assessment. Exchange of data and methods between the two project may be mutually beneficial; however high resolution data is not routinely dealt with in APHEIS.

SPAIN

Juan Ferrándiz
Virgilio Gomez
Pasqual Milvaques

Department of Statistics, University of Valencia, Spain

Ricardo Ocaña
Carmen Sanchez-Cantalejo
Antonio Daponte

Andalusian School of Public Health (EASP), Granada, Spain

Background

The Spanish version of the RIF has been completed and now it is operative in the Department of Epidemiology of the Valencian Regional Health Authority (DGSP hereafter as the acronym for Dirección General de Salud Pública) . We have resorted to the ORACLE and ARCVIEW licenses previously acquired by the University of Valencia and the DGSP.

We have developed two possible installation settings:

- **Standalone PC,** requiring the inclusion of ORACLE and ARCVIEW packages jointly with the RIF in a isolated PC running Windows 2000 OS, and
- **RIF satellites** of a central database, where many PCs are connected to the central ORACLE database held in a UNIX machine of the DGSP computer centre. Data is generated and updated in this central database while individual RIF packages are installed in different PCs according to staff specific tasks within the Department of Epidemiology.

The first possibility offers wide mobility so that the RIF can be run by DGSP staff not connected to the computer centre (there are some geographically dispersed services that could benefit from the RIF) . The main difficulty is the limited amount of data to be worked by a single PC.

The second alternative seems suitable for updating purposes as well as for disclosure restrictions on sensitive data. But it requires to coordinate activities from different services within the DGSP.
Health, demographic, socio-economic and environmental datasets, according to the list provided in the previous phase of the EUROHEIS project have been incorporated to this central database as well as in the standalone version of the RIF.

Gathering, debugging and conforming datasets to RIF formats have been time-consuming jobs and a library of ORACLE scripts has been elaborated for future use in similar tasks.

The AVENUE scripts provided by the UK partner have been adapted to specific formats of our datasets and some debugging of the AVENUE code has been necessary.

Intensive exchange between technical staff of both partners has been made through internet and practical improvements have flown in both directions.

Analysis with covariates

Most of the Spanish envisaged case-studies have to do with exposure to different contaminants that, although distributed with a clear geographical structure, are not reducible to the point-source analysis format . Instead of measuring exposure by distance to contamination origins we consider the level of these contaminants in the geographical units as a measure of exposure, much in the usual conception of ecological regression problems.

Thus we have added a new option within the RIF menu, offering this type of analysis besides those previously stated. Nevertheless we have kept mathematical complexity at the same level it was originally by performing this new analysis with the same computational techniques as those used with deprivation indexes. This strategy tries to keep computing times similar to previous values while avoiding substantial modifications of original RIF scripts.
The main steps of this new analysis are:

- Compare maps of disease and covariate
- Perform standardization of rates taking levels of the covariate as an added criterion to stratify the population (as we do with a deprivation index)

In this way we can see how filtering the effect of this covariate changes the resulting rates or maps.

Detailed description of instructions to perform this kind of analysis can be found in the attached technical report entitled 'Covariate Study' by Virgilio Gómez (Appendix 3).

This additional option to the basic RIF has been incorporated into the DGSP installed version. It is offered to the remaining partners if they consider it worth enough for their case-studies.

Missing data imputation

Some datasets, in particular those related to environmental data, present missing entries making their use in future studies difficult. The statistical models accounting for these nonexistent data are too complex to be incorporated in the RIF. First, because these models require computer-intensive statistical techniques and expert tuning of the computing process that are beyond the current capabilities of a 'rapid' tool like the RIF. Second because these complex models have to be tailored to the particular purpose of each study.

The traditional solution to this problem is the imputation of missing data. That means the prediction of lacking observations by fitting a good model to the observed data.
Future studies will take predicted data as if they had been observed and will apply standard techniques of analysis.

The problem remains of how the ignored uncertainty about the true values of missing data could affect the conclusions of the analyses. This is a topic still under current research and we have applied modern procedures to perform this task.

In particular we have fitted spatio-temporal models to our datasets to consider all the relevant information provided by the temporal sequences of measures in each region.

We have compared different models and approaches and we have used the best model in each particular dataset. The imputed values have been incorporated into the RIF in order to complete the registered series of data.

The comparative studies made on our datasets in order to choose the best imputation methods have been presented in the Seventh Valencia International Meeting on Bayesian Statistics and 2002 ISBA International Meeting held in the Canary Islands from June 1 to June 6, 2002, under the title 'Spatiotemporal multiple imputation in environmental studies' by C. Abellán, J. Ferrándiz and A. López.

Deprivation index

The necessary contacts with the Irish partner, leader of the development of deprivation indexes within the EUROHEIS project, have been established. Appropriate health and demographic data sets have been elaborated and initial steps of the analysis have been undertaken.

Case studies and preliminary checking

All along the process of RIF building, but most importantly in the final steps, we have being working with some case studies presented in previous phases of the EUROHEIS project. In particular we have considered datasets of:

- Cardiovascular mortality in Comunidad Valenciana and quality of drinking water (calcium and magnesium contents)
- Cancer mortality (prostate, bladder, colon) in Comunidad Valenciana and nitrate concentration in drinking water

We have performed disease mapping and covariate analysis on these problems in order to verify that everything runs smoothly.

A new case-study has emerged from the recent public alarm on air pollution in Alcora due to high concentration of tail industries. The DGSP is undertaking epidemiologic studies and projects to gather air-pollution measurements.

All these problems will be explored further in Phase III of the EUROHEIS project.

Dissemination

Spanish user's manual

This is an important tool in order to extend the use of the RIF among DGSP epidemiologists. We need a friendly manual explaining the purpose of every kind of analysis, the way it is performed within the RIF environment and how to interpret the resulting tables and maps. Clear writing and appropriate images are crucial.

We have written a first version in Spanish to be used by DGSP epidemiologists in the next phase of the EUROHEIS project . We expect to improve it by incorporating their criticisms and recommendations.

EUROHEIS presentation in professional societies meetings

We have pesented EUROHEIS project in some meetings of professional societies. In particular we can cite

Zurriaga O., Vanaclocha H., Abellán J.J., Martínez-Beneito M.A., Melchor I., Ferrándiz J., López A., Gómez V., Sanmartín P., Calabuig J., Ballester F., Pérez-Hoyos S., Daponte A., Ocaña R., Gutiérrez P. "Euroheis: Un proyecto Europeo de sistema de información geográfica, sanitaria y medioambiental para la valoración de exposición y riesgo y cartografía de enfermedades". XIX Reunión Científica de la Sociedad Española de Epidemiología}, Murcia, 18th-19th October 2001.

Ferrándiz J., López A., Milvaques P., Gómez-Rubio V. “Sistema de información geográfica en análisis medioambientales en salud pública”. Conferencia Internacional de Estadística en Estudios Medioambientales, 21st-23rd November 2001.

SWEDEN

Christina Reuterwall

Research and development Unit
Jämtland County Council
P.O. Box 602
S-832 23 Frösön
SWEDEN

Niklas Hammar

Institutet för miljömedicin
(IMM)
at Karolinska Institutet
Stockholm
SWEDEN

Implementation

The Stockholm RIF was an early adopter of the UK RIF. Thus, it also served as a trial and a test of future implementation strategies. The implementation has added to the library of Avenue scripts (programs that makes the system run and present data in a desired manner) and has identified special items to observe when transferring the RIF to new environments (in other countries). These experiences will be documented separately.

The implementation work during Phase II of the EUROHEIS project comprised the following tasks:

- add nominator (disease and mortality) data, denominator (population counts) data, and smoking and overweight class data for an extended period of time,
- account for revised geographical borders over time with reference to data aggregation,
- account for different revisions of the ICD (International Classification of Diseases) code,
- create a suitable socio economic index for the Stockholm county population and geo reference it,
- create data sets and shape files for air pollution exposures,
- install Oracle software, creating data tables and loading data sets into Oracle,
- install ArcView software and the shape files needed for using the RIF with the Stockholm county data,
- install the English RIF, adapt the system to the Stockholm prerequisites and test it.

Disease and mortality information, as well as socioeconomic exposure information is available on ‘base unit’ level.

Adding disease and mortality data

Age and sex specific data on morbidity, mortality, and population numbers has been retrieved and compiled annually and covers presently the period 1990-1999/2000. This information forms the basis for the following disease and mortality information:

- cancer incidences for 17 diagnose groups;
- mortality rates for 27 diagnose groups;
- hospitalisation rates for 35 diagnose groups.

Revised geographical borders

Political and administrative decisions have resulted in several changes of the borders of the geographic units of interest of the health information system. Look-up tables had to be constructed to allow for common geocoding across the time span 1990-2000.

Different revisions of ICD coding

Two different ICD coding revisions have been in effect during the time span: ICD-9 and ICD-10. This had to be accounted for to allow computation of moving averages and compilation of time trends for certain diagnose groups.

Socioeconomic characteristics

The availability of socioeconomic indicators suitable for use with health information system has been explored. This concerns indicators of special interest to the EUROHEIS project as basis for commonly defined indices of e.g., social deprivation. It also concerns indicators of special interest for Swedish/Stockholm settings. A modified version of the Townsend index has been developed for the Stockholm county. It is based on information collected in censuses and calculated per ‘base unit’.

Air pollution data

Information on NOx levels is retrieved from computer simulations using NOx concentrations measured at certain geographic points together with dispersion models. NOx information for 1995 is calculated for squares of different sizes within the Stockholm county; the size of the squares varies from 100 m * 100 m in the central part of the urban area to 1 km * 1 km in the rural parts of the county.

Shape files

The shape files contain map information, such as borders between base units, roads with different traffic intensities, lakes and rivers, electric power lines, etc. This is necessary background information for the disease mapping. Shape files, adapted to the ‘base units’ and the range of geographic squares used with air pollution data in the Stockholm RIF, had to be developed.

Case study

The Swedish partner will test and illustrate the capability of the implemented RIF as a tool for disease mapping and description of disease occurrence. The occurrence of acute myocardial infarction (AMI) in relation to social and socioeconomic determinants in the Stockholm region will be studied, and risk estimates obtained from this ecological (disease mapping) approach will be compared with results, based on individual data, obtained from a large case-control study. The study base of this study will consist of the population of Stockholm County during 1990-1999. New cases of acute myocardial infarction in the study population during this period will be identified using information from national registers of hospital discharges and deaths. This information will be combined using the unique Swedish personal identification number in accordance with previously developed and evaluated methods. Essentially, all new hospital admissions or death occurring within 28 days will be considered to reflect the same acute myocardial infarction episode. Swedish validation studies have suggested that this register based method of case identification has a high sensitivity and a high positive predictive value for definite acute myocardial infarction. For all cases information about place of residence at the time of disease onset, in terms of a small geographical area (base area), will be obtained through linkage to national population registers. Socioeconomic characteristics of the base area will be obtained using the Swedish 1990 census. These characteristics will be classified in accordance with the Townsend index to determine level of socioeconomic deprivation in a base area. The association between socioeconomic deprivation and incidence of acute myocardial infarction at the aggregate area level in Stockholm County will then be studied. This will be a collaborative effort between the Swedish and the UK partners. Information on various environmental factors (such as ambient air pollution or traffic noise) may also be geo-referenced and added to the database. The usefulness of the system in health planning and health improving strategies will be assessed.

UNITED KINGDOM

Dr Paul Aylin
Dr Lars Jarup

Small Area Health Statistics Unit
Dept. Epidemiology and Public Health
Imperial College
London
UK

Implementation and dissemination

The UK partner has been further developing the Rapid Inquiry Facility (RIF) to aid its implementation in several of the partner countries. The scripts have been rewritten in light of feedback from our partners. The system has been tested and finalised ready for roll-out. Documentation, including the data and hardware specifications has been completed and we have provided assistance (including site visits) to partner countries where required. See Appendix 1 for full documentation.

The UK partner has taken the lead on the development of the EUROHEIS website, which includes details of the project, project partners, interim reports, and a demonstration of the RIF. A leaflet summarising the EUROHEIS project has also been produced.

Case study

Title

Investigation of cancer incidence in areas exposed to high levels of bromate in East and West Hertfordshire

Background

Recent WHO guidelines have suggested a standard of no more than 25 mg/litre of bromates in drinking water, with a future EU standard of 10 mg/litre. In preparation for these standards, during routine monitoring, Three Valleys Water company discovered high levels of bromate at >150 mg/litre in water supplies around the Hatfield area and concentrations as high as 3,000 mg/litre in several private boreholes at Sandridge. This contamination is likely to have come from a chemical plant in Sandridge that operated from 1958 to 1972, manufacturing sodium and potassium bromates. There is currently no epidemiological data to suggest that potassium bromate is a human carcinogen. Most cases of human poisoning from bromate are due to the accidental or intentional ingestion of home permanent wave solutions, which contain 2-10% bromate. There is evidence in experimental animals of the carcinogenicity of potassium bromate.

Aims

To investigate the incidence and relative risk of selected cancers in the affected area in East and West Herts.

Design

Ecological study using the RIF.

Population
Population within contaminated water supply boundaries. Cancer incidence between 1975 and 1999. Diagnoses to include all cancers, malignant neoplasm of the peritoneum (ICD8&9 1580-9), malignant neoplasm of the thyroid gland (ICD8&9 193), malignant neoplasm of the kidney (ICD8&9 189), Malignant neoplasm of bladder (ICD8&9188).

Methods
Obtain digital boundaries of affected water supply from Three Valleys Water. Extract list of EDs with population centroids within boundaries. Group water supply areas into those with bromate levels >10mg/litre before 1991 and ascertain number of selected cancer diagnoses within these EDs from routine national cancer registrations. Using the Rapid Inquiry Facility, calculate incidence ratios standardized by age, sex and Carstairs quintile using government region as standard population. Document impact of report with local authorities.

Acknowledgements
The Small Area Health Statistics Unit is funded by a grant from the Department of Health, Department of the Environment, Transport and The Regions, Health and Safety Executive, Scottish Executive, National Assembly for Wales and Northern Ireland Assembly. The views expressed in this publication are those of the authors and not necessarily of the funding departments. This work is based on data provided with the support of the ESRC and JISC and uses census and boundary data, which are copyright of the Crown, the Post Office and the ED-LINE consortium.

MEETINGS AND DISSEMINATION

Meetings

January 2001: Department of Epidemiology, Imperial College, London, UK

April 2001: NATO Advanced Research Workshop GIS for Emergency Preparedness and Health Risk Reduction Budapest, Hungary
NATO Science Programme 2001

September 2001: ISEE Conference 2001, Garmisch-Partenkirchen, Germany
http://193.97.204.182/isee2001/abs/sheets/ABS00315.HTM

September 2001: APHEIS-EUROHEIS workshop, Garmisch-Partenkirchen, Germany
See appendix 4

April 2002: Escuela Andaluza de Salud Pública (EASP), Granada, Spain

Publication

Geographical Information Systems (GIS) for Emergency Preparedness and Health Risk Reduction.
Published by Kluwer Academic Publishers, PO Box 17, 3300 AA Dordrecht, The Netherlands

EUROHEIS website

http://www.med.ic.ac.uk/divisions/60/euroheis/homepage.htm

Appendix 1. RIF documentation

Appendix 2. Details of the Bayesian models and model fitting

Appendix 3. Covariate study

Appendix 4. Minutes of the APHEIS/EUROHEIS workshop, Garmisch-Partenkirchen, 6 September 2001

All appendices are attached as separate documents