Environmental public health: A disease mapping and risk analysis tool

Beale L, Abellan J, Andersson M, Hambly P, Hodgson S, Jarup L

Small Area Health Statistics Unit,
Department of Epidemiology and Public Health,
Imperial College London,
London, W2 1PG

Abstract
Within the framework of the Centers for Disease and Control Prevention Environmental Public Health Tracking Program (EPHT) the Small Area Health Statistics Unit (SAHSU) has further developed their Rapid Inquiry Facility. The RIF is embedded in ArcGIS and can link environmental and health databases using open database connectivity to enable rapid handling of environmental public health concerns.

The RIF produces rates and relative risks for any specified disease endpoint in relation to environmental sources of pollution adjusted by age and sex and, for any number of other covariates. The RIF can also produce disease maps at a small area level, using Bayesian smoothing techniques to minimize problems of sparse case data in small areas.

Users can export results in a number of formats including word reports, data tables and as maps that include the visualization of uncertainty. Further analysis can be done through direct linkage to SaTScan and WinBUGS.

Introduction
Development of the Rapid Inquiry Facility (RIF) was first reported in the Journal of Pubic Health Medicine, September 1999. It was described as a national facility for small area disease mapping and rapid initial assessment of apparent disease clusters around a point source (Aylin P, Maheswaran R et al, 1999). Recent collaboration between the small area health statistics unit (SAHSU), Imperial College with the Centers for Disease Control and Prevention (CDC) has allowed development and significant enhancement of the RIF software for use in CDC's National Environmental Public Health Tracking (EPHT) Network. The goal of this collaboration is to increase the functionality and versatility of the RIF for evaluating spatial relationships between disease and environmental hazards in the National EPHT Network.

The RIF was initially designed for use in the Small Area Health Statistics Unit (SAHSU). SAHSU was established in 1987 following a recommendation of the Black Inquiry into the incidence of leukaemia in children and young adults near the Sellafield nuclear plant. The main aim of SAHSU was to assess the risk to the health of the population from environmental factors, with an emphasis on the use and interpretation of routine health statistics (the Unit holds national mortality, cancer, congenital anomaly, birth, stillbirth and hospital admission data, with the postcode of residence to locate individual events at 10-100 m resolution). The RIF enabled SAHSU staff to rapidly analyse routinely collected health data in relation to environmental exposures in the UK, with options for assessing risk in populations living around putative hazardous sources, or by mapping disease across an area.

The original UK RIF could generate age, gender and socio-economic standardised rates and risks for any given health outcome, for specified age and year ranges, for any geographical area in the UK, relative to a national or regional reference population. To aid interpretation the RIF would also display contextual geographical information, in additional, for the hazard analysis function, graphs comparing the age, gender and socio-economic make up of the study and comparison population were also produced, and for the disease mapping function, smoothed small area risk maps were available in addition to the crude and age/gender standardised risks. This tool was used by SAHSU staff, working in conjunction with local public health specialists in England and Wales, to rapidly assess numerous potential disease clusters in relation to pollution sources.

Methods
The investigation of possible health effects around point sources of environmental pollution have traditionally been costly and time consuming since health and population data relevant to the area under study would need to be assembled and analysed ad hoc. The RIF can automate this process by rapidly retrieving and analysing the relevant data, provided several criteria are met:

1. The availability of accurate health event data, which has a high level of ascertainment (to minimise errors in estimation of local rates), and which is located geographically to a place of residence or small geographical area.
2. Population data (e.g. from a national census) by small geographical area, and by age and gender, to allow calculation of disease rates.
3. Geographical links between the health events (numerator) to the population data (denominator).

To use the RIF, the user must select a 'study population', for example defining the population living within a certain distance of a point source of pollution, or the populations of administrative areas falling within the jurisdiction of a health authority. The RIF retrieves all specified health events occurring over a selected time period that fall within this study area. The population living within the selected study area is also retrieved to act as the denominator. The expected number of cases is calculated by applying regional or national age and gender specific rates (calculated within the database) to the study population. The standardisation procedure can be extended to include stratification by other variables such as socio-economic status, income or ethnicity. In this way the RIF calculates indirectly standardised mortality, morbidity or incidence ratios (SMRs). Directly standardised rates are also calculated by applying the disease rates in the study population to the population of a comparison region.

Whilst these basic functions of the RIF have formed the core of this tool since its original development, the new RIF being developed in collaboration with the CDC has much increased flexibility and additional functionality.

The new RIF provides an extension to ESRI® ArcGIS (v9.0 and higher) functions, and uses both database and GIS technologies. By being embedded within ArcGIS, RIF users have access to all the functionality that ArcGIS offers, thus the RIF provides epidemiologists with a tool to take advantage of the many of the functions that ArcGIS offers, without requiring an in-depth knowledge of GIS.

The RIF uses health and population data (and optionally covariate data) that is stored in a database. Traditionally, within SAHSU, that is been an Oracle database, due to size of the data held. Development of the RIF, has used ODBC technology to link ArcGIS to the database, so that other database types can be used, for example ACCESS. All calculations are carried out in the databases, using SQL, whilst user input and results are handled by ArcGIS.

RIF Studies

The RIF allows two different types of study to be carried out namely, disease mapping and risk analysis. There are four steps in setting up a study that include defining the scale of analysis, defining of the study area and the comparison area, and finally the selection of the investigation details.

The risk exposure analysis allows the user to calculate both directly and indirectly standardised rates within user defined distance bands around one or more user defined point or area sources. Furthermore, shapefiles, such as outputs from dispersion models, can be used to define source areas. Since the RIF allows all the ArcGIS functions to be accessed a variety of import options, and data selection methods can be used. The populations at risk are selected spatially, from the chosen census, but using a refining technique that employs the area centroids. These can be defined using a shapefile, for example, of population-weighted centroids or alternatively where such data does not exist, geographical centroids are dynamically calculated.

The disease mapping functionality of the RIF allowed a user to produce maps of directly standardised and indirectly standardised rate ratios. It also allows smoothing of indirectly standardised rate ratios via empirical Bayesian estimation. Using empirical Bayes Estimation to smooth the raw relative risks attempts to account for sampling variability in the observed data. Unstable risk estimates can occur when investigating health risks in small areas where small populations have small number of expected and observed events. This can result in misleading risk maps, especially if the areas with the smallest populations cover large regions (e.g. rural areas), as these areas with the least stable risk estimates can dominate a map.

A number of export options are available in the RIF. Results are displayed in ArcMap but additionally reports can be automatically generated. The new reporting system generates XML structured data (text) for interchange between computer applications such as MS Word. The RIF uses these XML data and a MS Word template to seamlessly generate a report. The templates have key words that are replaced with

text, tables, charts or maps from the data in the XML. Furthermore, the MS Word templates can be easily modified to change formatting, add or remove text, insert logos and so on.

The RIF can also export data for further analysis in other (statistical) software packages such as EXCEL, WinBUGS and SatScan. Recent developments have included direct linkage to both SatScan and WinBUGS which not only allows seamless transfer between the packages but also automatically displays the results from these external packages in ArcGIS. This has meant that some uncertainty can be displayed, simultaneously with the results, which can aid interpretation.

Conclusion

The RIF is able to rapidly link environmental and health databases, and is thus a powerful tool for evaluation of spatial relationships between disease and environmental hazards. It allows researchers and public health professionals to calculate rates and risks by small area easily and rapidly, offering a tremendous time advantage over traditional manual data extraction and calculation. However, while the RIF can dramatically speed up data analysis and add to the interpretation of a potential disease cluster, it cannot overcome the numerous issues inherent in small area spatial epidemiology, nor can the RIF produce meaningful output if the exposure, health or population data are of poor quality. There are several issues to consider when undertaking any risk analysis or disease mapping exercise, whether this is carried out via a traditional manual approach, or through the RIF.

References

Aylin P, Maheswaran R, Wakefield J, Cockings S, Jarup L, Arnold R, Wheeler G, Elliott P, 1999, A national facility for small area disease mapping and rapid initial assessment of apparent disease clusters around a point source: the UK Small Area Health Statistics Unit, Journal of Public Health Medicine 21 (3), pp 289-298