

Appendix 1

EUROCIM software and databases

Guidelines for use of EUROCIM data

The EUROCIM package comprises:

- (i) the databases of cancer incidence and mortality data;
- (ii) the data analysis software.

The package is available only to registries which contribute data to the EUROCIM project.

The following guidelines for the use of EUROCIM should be adopted:

For contributing cancer registries

Registries wishing to compare their own data with those of other registries for annual reports etc., at the level of the main site categories (i.e. those in the main tables of *Cancer Incidence in Five Continents*), should be able to do so as long as they acknowledge the original data providers and provide them with copies of the resulting publications. For more detailed analyses (e.g. using the epidemiological entities) which might lead to a scientific publication, collaborative effort and shared authorship of publications are strongly encouraged. At the very least, the contributing cancer registries must be consulted before any submission for publication, to allow local experts to comment and advise on the data, analysis and interpretation, and to discuss authorship and/or acknowledgement issues.

For other users of cancer registry data

Researchers who are not part of a cancer registry organization which is participating in the EUROCIM project are advised to approach a local cancer registry. Registries should be free to use the EUROCIM data in collaboration with external researchers as long as they can ensure that the guidelines above are respected.

General information about EUROCIM

EUROCIM is a facility developed by the European Network of Cancer Registries. The current version (4.0) comprises a comprehensive database of cancer incidence and mortality, together with analytical software. Information about participating registries, and contact details, are available via the ENCR website (www.encl.com.fr).

The flexible facilities enable the user to construct Working Datasets comprising populations covered by selected cancer registries, years, age/sex groups, and cancers of interest. These can be investigated using a suite of statistical tools within EUROCIM. There is also a comprehensive, context-related Help system.

Output from EUROCIM sessions can be produced in the form of reports or graphs, and a comprehensive range of statistical functions is provided for analysis. In addition, export facilities are provided to allow EUROCIM data and outputs to be used by other applications.

The EUROCIM databases

Incidence data are obtained from European population-based cancer registries. The data are classified according to either ICD-10 codes, or combinations of ICD-O topography and morphology codes. Cancer registries in Europe use a variety of classifications to record cancer incidence, so the data in EUROCIM have been converted to ensure consistency. The original coding systems used by the registries are listed in the Help section of the corresponding database.

Mortality data are also available in respect of the participating registries. These are based on official national statistics and the WHO Mortality Databank, and are classified to ICD codes.

The cancer registries included in the EUROCIM database and the years for which their data are available are listed separately for incidence and mortality.

Creating a new population and/or entities

EUROCIM permits users to group populations and cancers of interest from the selected database and to save them as user-defined populations and user-defined cancer groups respectively (Figures 1 and 2).

Standard populations

EUROCIM provides the World and European standard populations for age-standardization purposes, and also allows users to define their own standard populations.

Working Dataset

A Working Dataset is a user-defined subset of the database. It comprises selected populations served by specified cancer registries, for specified years, age/sex groups and cancers of interest. The user can create a new Working Dataset, or open an existing Working Dataset defined and saved in a previous session. It is also possible to modify a Working Dataset.

The Working Dataset, once defined, can be used for generating models for statistical analysis, and output can be generated in the form of reports and graphs.

Data in a Working Dataset is displayed in a grid in which the columns represent cancers by sex, and the rows represent populations/periods for the age-range selected. The grid appearance can be configured by the user using the *Edit* menu. The grid displays either Crude or European Standardized Rates, depending on the user's specification (Fig. 3).

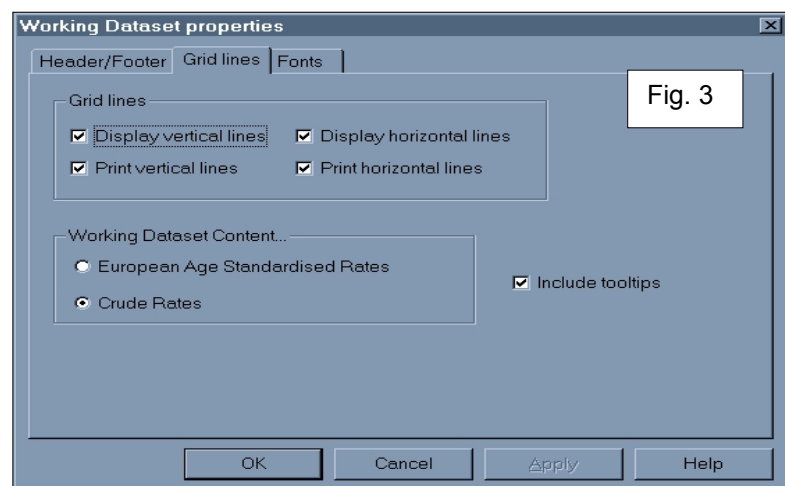
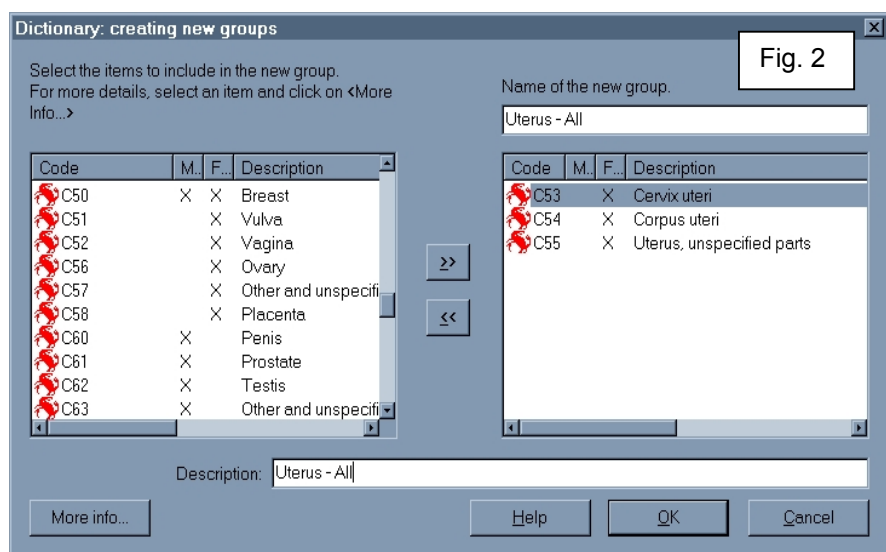
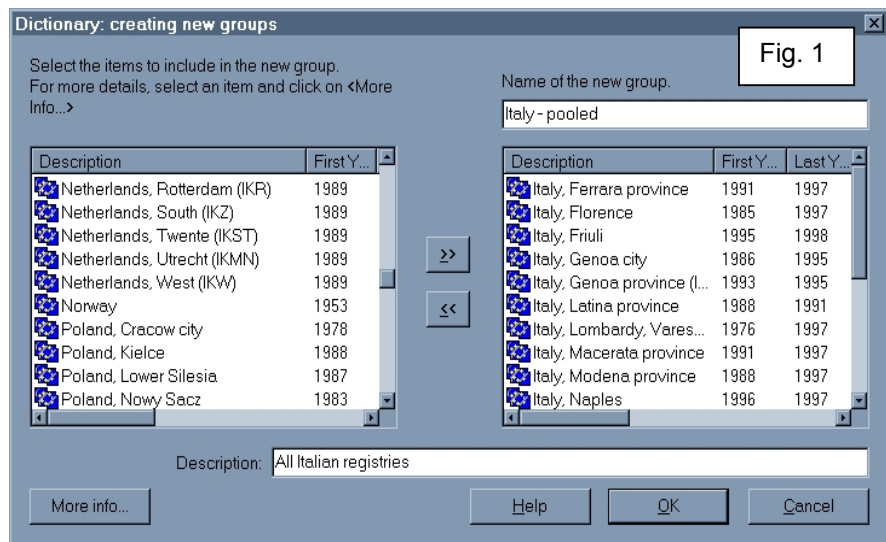
Report facilities

EUROCIM includes facilities for generating reports from data in the current Working Dataset. A Report Wizard is provided to facilitate this. Generated reports may include age-specific and summary rates, or information on a statistical model.

The following statistics can be included in reports:

- *Observed numerators*
- *Expected numerators*
- *Residual numerators*
- *Observed rates*
- *Expected rates*
- *Residual rates*

- *Denominators*
- *Age-specific parameters*
- *Age-standardized rates*
- *Age-truncated rates*
- *Cumulative rates*
- *Internal SMRs*
- *Maximum likelihood estimates*



	A	B	C	D	E
1					1996
2	US_LOWER				
3		1906			
4			male		
5				Age Standardised Rate - World	1.10 [0.44, 1.75]
6				Truncated Rate - European [0-85+]	1.48 [0.60, 2.35]
7				Cumulative Rate - [0-84]	0.13
8				Crude Rate	1.47
9				Internal SMR	1.55 [0.77, 2.77]
10			female		
11				Age Standardised Rate - World	0.27 [0.00, 0.58]
12				Truncated Rate - European [0-85+]	0.40 [0.00, 0.84]
13				Cumulative Rate - [0-84]	0.06
14				Crude Rate	0.51
15				Internal SMR	0.51 [0.14, 1.30]

Fig. 4

It should be noted that the following statistical functions require the creation of a log-linear model:

- *Expected numerators*
- *Residual numerators*
- *Expected rates*
- *Residual rates*
- *Age-specific parameters*
- *Maximum likelihood*

Figure 4 shows an example of output from a report.

Graph facilities

EUROCIM includes facilities for creating graphs from data in the current Working Dataset. It is possible to create graphs based on either age-specific or summary rates, as well as to display fitted values from the current model if generated. A *Graph Wizard* is provided to facilitate this. The statistics that can be displayed on the graphs are:

- *Observed numerators*
- *Expected numerators*
- *Observed rates*
- *Expected rates*

By default all curves defined in the Working Dataset are included in the graph. One can override this and select only the curves desired.

Age-period-cohort (time trends) module

It is possible to formally examine age, period and cohort trends in cancer rates over time using the command *APCView*, which fits age-period-cohort models to the selected population's incidence or mortality rates. A *Time Trends* working dataset by definition must include at least 15 years of incidence or

mortality data from one population (with no years missing). In addition, only one cancer site and sex can be analysed at one time.

The active Working Dataset displays the selected data. To examine age, period and cohort effects click on 'APC View' from the 'Working Dataset' menu. There are now three windows which display information on the fit of

each of the models; the parameters obtained; and to help with interpretation, a graphical display of the parameters.

Methodological aspects of the APC modelling

The analysis of the temporal variation of cancer incidence and mortality rates is one of the main concerns of descriptive epidemiology. While graphical displays of age-specific trend data are invaluable, they can be greatly enhanced by the use of statistical modelling by providing quantitative and comparable estimates of trend. These are based on objective criteria for choosing the best description of the data, and statistical tests to decide whether trends are real or random (Estève *et al.*, 1994) and consequently subjective graphical interpretations are avoided. The age-period-cohort model (APC) (Clayton & Schifflers, 1987a, b; Holford, 1983; Mason & Fienberg, 1985) provides such a summary of trend data, allowing a formal statistical examination of whether temporal trends are due to secular changes in risk (period effects) or changes in risk from generation to generation (birth cohort effects) in different populations. EUROCIM Version 4 provides a means to describe such trends in Europe, using the APCView option.

Modelling age, period and cohort simultaneously is not a simple task. The fitting of such effects mean that the resulting models suffer from the problem of identifiability. For A age groups, if we denote a_i as the age effect in age group i , p_j the effect of j th period and c_k the birth cohort effect, then age, period and cohort effects are all linearly dependent on one another as $k=A-i-j$. Thus there are an infinite number of possible solutions to the

“full” APC model (the model containing age, period and cohort effects) because the separation of the linear effects of the three parameters is impossible.

There are several subsets of the full model for which the parameters are estimable. The adequate fit to the data of an age-period model suggests that secular trends may be of importance. An exposure to a sudden intervention which affects the risk of each age group equally would result in such a model – changes in disease coding over time (which affect all age groups) is one such example. In contrast, the age-cohort model will provide a good fit to the observed data if some intervention affects the age groups in different ways; such an effect is a special form of the age-period interaction. The multistage theory of carcinogenesis (Day & Brown, 1980) suggests that for some cancers (given their long latency period), changes in exposures related to the underlying risk should show up more clearly in birth cohorts (Cuzick, 1990).

Clayton and Schifflers (1987a) introduced the term “drift” to describe a model for which age-period and age-cohort parameters fit the data equally well. The model implies the same linear change in the logarithm of the rates over time in each age group and thus such a model serves as an estimate of the rate of change of the regular trend. It can be shown that period-drift and cohort-drift models lead to identical fitted rates. In assessing the best-fitting model, Clayton and Schifflers (1987a) recommend that the age-drift model should be fitted after the model of no temporal trend (the age model) in the hierarchical fitting process. This convention is used in APCView. The model-fitting process continues with the fitting of age-period and age-cohort models (which include drift). Such time trend analyses are however not so absolute as to allow one to attribute incidence or mortality to purely secular or generation effects. Any gradual linear effect can be attributed to either factor, given that only changes of a non-linear nature are identifiable—thus their interpretation should always be on the cautious side.

When neither age-drift, age-period or age-cohort terms give an adequate explanation of the data, the full APC model in which age, period and cohort effects are included, needs to be considered. As mentioned above however, the algebraic relationship between the effects is such that it is impossible to obtain a unique solution of their linear effects (there are an

infinite number of them) without resorting to some (often arbitrary) constraint (Clayton & Schifflers, 1987b). The drift component defined above cannot be attributed to period or cohorts effects and thus the APC model is a mixture of age, drift, non-linear period and non-linear cohort effects. There has been a considerable body of work on how one can then proceed in view of this identifiability problem in the APC model.

The main differences in the methodologies are in the way they differentiate between linear and non-linear effects (Holford, 1992). In choosing a method for which to describe secular trends it seems appropriate to avoid imposing constraints without a biological basis for doing so. The “second differences” method of Clayton and Schifflers (1987b) has been shown to produce estimates on the conservative side, and is not particularly easy to interpret. Nevertheless, it is preferable to other methods, as it does not make further assumptions (McNally *et al.*, 1997) or require an additional arbitrary mathematical constraint to overcome identifiability. In addition, the risk of over-interpretation of the resulting parameters (perhaps more serious than under-interpretation) is more likely to be avoided using this method. It is therefore this method which has been used in EUROCIM's *APCView* to provide a unique set of parameters representing the full APC model.

References

- Clayton, D. & Schifflers, E. (1987a) Models for temporal variation in cancer rates. I: Age-period and age-cohort models. *Stat. Med.*, **6**, 449–467
- Clayton, D. & Schifflers, E. (1987b) Models for temporal variation in cancer rates. II: Age-period-cohort models. *Stat. Med.*, **6**, 469–481
- Cuzick, J. (1990) International time trends for multiple myeloma. *Ann. N. Y. Acad. Sci.*, **609**, 205–214
- Day, N.E. & Brown, C.C. (1980) Multistage models and primary prevention of cancer. *J. Natl Cancer Inst.*, **64**, 977–989
- Estève, J., Benhamou, E. & Raymond, L. (1994) *Statistical Methods in Cancer Research*, Volume IV, *Descriptive Epidemiology* (IARC Scientific Publications No. 128), Lyon, IARC
- Holford, T.R. (1983) The estimation of age, period and cohort effects for vital rates. *Biometrics*, **39**, 311–324
- Mason, W.M. & Fienberg, S.E. (1985) *Cohort Analysis in Social Research: Beyond the Identification Problem*, New York, Springer-Verlag
- McNally, R.J., Alexander, F.E., Staines, A. & Cartwright, R.A. (1997) A comparison of three methods of analysis for age-period-cohort models with application to incidence data on non-Hodgkin's lymphoma. *Int. J. Epidemiol.*, **26**, 32–46