EUROCARE-4

PROTOCOL FOR SPECIFIC PROJECT ON PERIOD ANALYSIS

4 November 2004

1. RATIONALE AND AIMS

Period survival analysis[1-7] is a new, but already established, method allowing the incorporation of the most recent follow-up information available into the analysis of survival for the most recently diagnosed cases. It enables an estimation to be made of (say) five-year survival before all patients have been followed up for five years.

Period survival analysis has been extensively validated on Finnish Cancer Registry data[3]. For almost all cancer sites, it has been shown to provide a closer estimate of the actual survival of the most recently diagnosed patients than classical cohort survival analysis. However, period survival indicators are intrinsically more affected by possible incompleteness of registration and follow-up than cohort-based indicators. A certain proportion of incident cases only become known to the registry some years after the date of diagnosis. The proportion of late registrations will depend on which information sources and data collection techniques are used by the registry, and this proportion may therefore differ substantially between cancer registries. If late registrations are not random, and selectively involve patients who have a good (or bad) prognosis, then period survival estimates may be subject to a registry-specific bias, which in turn may invalidate survival comparisons. Furthermore, some death certificates may arrive late at the cancer registry, with the result that some patients who actually died before the end of follow-up may erroneously be reported in that registry's data as still being alive.

This project therefore has the following main objectives:

- To study the possible lack of quality and completeness in the most recent cancer registry data, and the effect of any differences in quality on period survival estimators
- To provide systematic, large-scale application of period survival analysis to European cancer registry data. This objective is conditional on the feasibility of the approach having been demonstrated (previous objective), and the conclusions from those analyses will be taken into account in the presentation and discussion of results.
- To update, by means of period survival analysis, the comparisons of European survival data with data from the USA, in the framework of the CONCORD study.

For confidentiality, security and publication policy, please refer to the EUROCARE-4 main study protocol. In order to avoid unnecessary duplication of work in sending, preparing and analysing the data files, the same record structures and data formats are required for this project as those requested for the main EUROCARE-4 study. The only difference is the more recent incidence period requested for the period survival analysis. Thus, data of registries participating in the specific study on period survival will be automatically included also in the general EUROCARE-4 survival project

2. CRITERIA FOR INCLUSION

Registries are requested to send updated data for the whole incidence period 1978-2002, if possible with the same identification code(s) used in EUROCARE-3. This will enable the Data Analysis Centre to avoid unnecessary repetition of quality checks, and thus to minimise requests to participating cancer registries for correction or confirmation of data in records with unusual codes.

All cancer sites are to be included in the data. Data for ***all tumours*** should be included, including death-certificate-only cases (DCOs), cases discovered only at autopsy and cases lost to follow-up. Both histologically verified and unverified cases must be included. Anatomic tumour location (site) is to be coded according to the International Classification of Diseases: both 9th and 10th revisions are acceptable[8,9]. Morphology is to be coded according to the International Classification of Diseases for Oncology (ICD-O). Preferably 2nd , but also 1st and 3rd editions are accepted[10,11,12].

For bladder neoplasms only, benign and *in situ* transitional cell papilloma/carcinoma will also be included in the survival analyses. For all other anatomic sites, *in situ* and benign tumours (as identified by the ICD-O behaviour code) will be included in the data base but *not* in survival analyses. The frequency of these tumours will assist in the interpretation of survival differences between registries as an indicator of the likely intensity of early diagnostic activity in the territory covered by the registry.

In the case of multiple malignant tumours, only the first malignant tumour in a person will be included in the basic survival analysis (non-melanoma skin cancer and *in situ* tumours, however, will be ignored as a first primary malignancy). Registries are requested to provide a separate file of tumours eliminated as the second or subsequent malignant tumour in a given person, in addition to the file(s) of first primary malignancies for inclusion in the main analyses. When possible, the same identification code should be given to all tumours that occurred in the same person, to enable linkages to be checked if necessary, or for cross-tabulation of first and second tumour frequencies in different registries as part of overall quality control.

The data structure for records in the file of second or subsequent tumours should be the same as for tumour records in the main survival data file.

3. DATA TO BE COLLECTED

Unique codes for country and registry will be assigned to every participating centre and placed automatically at the beginning of each individual record in the survival data file by the Data Analysis centre. Registry codes used in *Cancer Incidence in Five Continents* will be used, in order to facilitate the integration of EUROCARE data with other data bases.

Data should be sent by surface mail in any standard storage device (or by safe electronic transmission), with an accompanying letter describing the content.

This letter should contain at least:

- the name and address of the registry
- an explicit statement of agreement to participate in the period survival analysis project, signed by the Registry Director (or equivalent person)
- the name of the reference person in the registry who can be contacted for data checks
- the most recent date for which follow-up is considered by the registry to be complete
- details of ICD or other classifications used to code tumour site and morphology
- detailed explanation of any other non-standard codes used by the registry (e.g. stage, treatment)

The data will be stored and submitted to basic centralized data checks and analysis at the Istituto Superiore di Sanità in Rome, Italy.

Data should be sent to:

Dr Riccardo Capocaccia
Centro Nazionale di Epidemiologia, Sorveglianza e Promozione della Salute
(National Centre for Epidemiology, Surveillance and Health Promotion)
Istituto Superiore di Sanità
Viale Regina Elena, 299
I-00161 Roma, Italy

4. DATA ANALYSIS

4.1 Quality and completeness of recent cancer registry data

A special analysis of data quality and completeness will be carried out on the data from each registry. Possible incompleteness of follow-up due to late arrival of death certificates will be investigated by calculating very short-term (one year or less) survival ratios for patients with a lethal cancer diagnosed in the last year. These will be compared with the corresponding figures for patients diagnosed in previous years, for which follow-up to determine vital status can be assumed to be virtually complete.

The number of cases diagnosed in the last available year will be compared with the corresponding number for previous years. A linear trend will be fitted to the incidence data from each registry by join-point analysis[13], excluding the last available year. An estimate of completeness in the last available year will be given by the difference between the expected number of incident cases based on trend analysis and the corresponding observed number. Where completeness is estimated to be substantially less than 100%, the distribution of cases by selected key variables - age, site (including unspecified site), morphology (including non-specific or unspecified histologies), vital status, mode of diagnosis (including DCO) - will give some indication of the prognosis of missing cases.

A similar analysis can be carried out for registries in which we have two consecutive submissions of incidence data for the same year. For example, this is the position for registries in the US Surveillance, Epidemiology and End Results (SEER) programme, for which we now have 1999 incidence data included in both 2002 and 2003 public use data releases. Comparing the number of 1999 incident cases, and their distribution by age, site, vital status, stage and mode of diagnosis will provide a description of tumours that were not included in the 2002 data set but were subsequently included in the data released in 2003.

More detailed analyses should be possible with information on the date of registration of each tumour. This is a new variable in the EUROCARE project, requested for the first time in the EUROCARE-4 protocol. Let "*registration delay*" be the difference between the date of registration and the date of diagnosis. The objective of EUROCARE-4 is to analyse the survival of patients diagnosed up to the year 2002 and for whom cancer registration and follow-up data can be provided by the registry within the first quarter of 2005. In this context, therefore, 'late' registrations may be defined as those with a *registration delay* of two or more years. For registries able to provide the date of registration for all cases, the proportion of late registrations can be estimated from the historical data. This proportion can be also analysed by the extent of *registration delay* (2, 3, 4 years, etc.) and by other variables such as the year of diagnosis, age, sex and site. Survival rates for late registrations will be compared with those for other cases, i.e. those included in the data within two years of the date of diagnosis. A model-based approaches[14,15] could also be applied to the data to estimate the trend in completeness by registry for each of the main cancer sites.

This analysis will be carried out on the whole data set provided by each registry, i.e. not limited to the most recent years of incidence for which data are available. Therefore, *cancer registries that are not able to provide incidence data for the period 2000-2002 can still participate in this component of the study*.

4.2 Survival analysis

Data for all cases collected by participating cancer registries and diagnosed from 1978 to 2002 inclusive should be sent to the Data Analysis Centre in Rome, using the record structure indicated in Section 3. A fixed format record is requested for each tumour. Data files will be submitted to the same checks and procedures that have been developed and tested in the framework of previous EUROCARE studies and published on the EUROCARE website www.eurocare.it.

All data files should be received in Rome by 31 May 2005, in order for preparation of the data base for analysis to be completed before the end of 2005.

The principle of period survival analysis is simple. Only the most recent period of follow-up is considered for the analysis. In order to provide long-term survival estimates, all observations included in the analysis are left-truncated at the beginning of the period of interest and right-censored at the end of it. For example, if we are interested in estimating survival based on probabilities estimated from the follow-up (death, emigration …) data for our cancer patients recorded during the period 2000-2002, then survival experience during the first year after diagnosis is provided by patients diagnosed between 1999 and 2002. Survival experience in the second year after diagnosis is provided by patients diagnosed between 1998 and 2001, and so on, until survival experience in the tenth year after diagnosis, which is provided by patients diagnosed between 1990 and 1993.

Period survival analysis methods will be used to estimate one-, three-, five- and ten-year relative survival rates for incident cases diagnosed up to and including the year 2002, on the basis of the survival experience of all patients during the period 2000-2002. Trends in period survival will be estimated by examination of survival for patients diagnosed in successive three-year periods of the available incidence data.

Data will be presented by site, sex and registry. Aggregation of registries into geographical areas will be attempted, but it is difficult to anticipate the feasible degree of aggregation (country, European regions, Europe as a whole), since this will depend on the availability of data and the geographic homogeneity of survival rates.

4.3 Life tables of general population mortality

This project will use the same updated life tables collected as part of the main EUROCARE-4 study.

5. PUBLICATION OF RESULTS

A first paper will be written illustrating the results of the quality and completeness analyses, including an estimation of the possible effects of incompleteness on the period survival estimates.

The main paper will provide estimates of cancer survival in Europe for the main cancer sites using follow-up experience of registered cancer patients during the years 2000-2003. We plan to prepare both these papers within the year 2006. The final date of submission will be discussed with the EUROCARE Steering Committee for a full co-ordination with the other outputs of the project.

One or more papers will then be planned to present changes in period survival for the most interesting cancer sites. A more detailed proposal for publication of the analyses of trends in period survival will be prepared when the preliminary results become available.

REFERENCES

1. Brenner H, Gefeller O. An alternative approach to monitoring cancer patient survival. *Cancer* 1996; 2004-2010.
2. Brenner H, Gefeller O. Deriving more up-to-date estimates of long term patient survival. *J Clin Epidemiol* 1997;211-216.
3. Brenner H, Hakulinen T. Advanced detection of time trends in long-term cancer patient survival: experience from 50 years of cancer registration in Finland. *Am J Epidemiol* 2002;566-577.
4. Brenner H. Long-term survival rates of cancer patients achieved by the end of the 20th century: a period analysis. *Lancet* 2002;1131-1135.
5. Brenner H, Gefeller O, Hakulinen T. A computer program for period analysis of survival. *Eur J Cancer* 2002:38; 690-695.
6. Brenner H, Hakulinen T, Gefeller O. Computational realization of period analysis for monitoring cancer patient survival. *Epidemiology* 2002:13;611-612.
7. Brenner H, Gefeller O, Hakulinen T. Period analysis for up-to-date cancer survival data: theory, empirical evaluation, computational realisation and applications. *Eur J Cancer* 2004:40;326-335.
8. Manual of the International statistical classification of diseases, injuries, and causes of death. Ninth revision (ICD-9). World Health Organisation Geneva, 1977
9. International statistical classification of diseases, injuries, and Related Health problems. Tenth Revision (ICD-10).World Health Organisation Geneva, 1991.
10. International Classification of Diseases for Oncology. First edition. World Health Organisation, Geneva, 1976
11. Percy C, Van Holten V, Muir CS (eds.). International Classification of Diseases for Oncology. Second edition. World Health Organisation, Geneva, 1990.
12. Friz A, Percy C, Jack A, Shanmurgaratnam K, Sobin L, Parkin DM, Whelan S (eds.). International Classification of Dieseases for Oncology. Third edition. World Health Organisation, Geneva, 2000.
13. Lerman PM. Fitting segmented regression models by grid search. Applied Statistics 1980: 29;77-84.
14. Bullard J, Coleman MP, Robinson D, Lutz J-M, and Peto J. Completeness of cancer registration: a new method for routine use. BJC 82 (2000), 1111-1116.
15. Clegg LX, Feuer EJ, Midthune DN, Fay MP, and Hankey BF. Impact of reporting delay and reporting error on cancer incidence rates and trends. JNCI 94 (2002), 1537-1545.