Measuring and predicting cancer incidence, prevalence and survival in Finland

Timo Hakulinen and Karri Seppä
Finnish Cancer Registry- the Institute for Statistical and Epidemiological Cancer Research

ABSTRACT

The entire country of Finland has been covered by cancer registration since 1953. The registration is compulsory and based on law. On average, five notifications per case are received by the cancer registry from various sources incl. hospitals, laboratories and the central statistical office. Record linkage and follow-up are facilitated by the use of personal identification numbers in the population. The completeness of the cancer registration and the follow-up is secured and verified by linkages with the population register, other registers and study materials.

Because of the almost complete follow-up and the long-term existence of the cancer registry, not only the incidence, but also the prevalence of cancer can be directly estimated. A problem in this is, however, the decision of whether a single patient can be regarded cured or not. This cannot be directly inferred based on the data of the cancer registry, but relative survival analysis of the patients proves to be helpful in the estimation of proportions of cured patients. This analysis can also be used to determine a possible point of cure during the follow-up after which all living patients can be regarded cured. Before that point it is not possible to judge on the basis of the cancer registry data which patients are cured even though their proportion can be estimated.

The survival analysis figures give both the estimates on the net (cancer-specific) and the crude (total) probabilities of surviving by follow-up time as well as the probabilities of death by cause.

In Finland, in a population of about five million residents, about 220 000 persons have had cancer some time in their life. During the first five years after diagnosis, approximately one third of the patients succumb to their cancer and 10% to other causes of death. Predictions made for these and other quantities provide a firmer background for resource planning and cancer policy in the years to come.

THE FINNISH CANCER REGISTRY

Cancer registration in Finland was started by the Cancer Society of Finland in 1952 but the Society had already begun some pilot registration in the late 1930s. The first complete cancer register year was 1953. The Finnish Cancer Registry is the organisation the Cancer Society founded to take care of the database, the cancer register. Today, the Finnish Cancer Registry is still run by the Cancer Society, it has about 40 employees and 30% of the total costs are covered by the state (the cancer register database maintenance and statistics production costs). The Registry’s activities result in about 100 peer-reviewed publications per year.
THE CANCER REGISTER

The registration is compulsory and based on law. The Registry receives notifications from hospitals, laboratories and the central statistical office (the latter include death certificates where cancer has been mentioned and the mortality follow-up of the already registered patients). On average, five notifications per case are received. Each of the notifications contains the personal identification number of the patient, which makes the record-linkages reliable and efficient. The follow-up is secured also with respect to emigrations by regular linkages to the national population register. The completeness is verified by linkages with other registers (e.g., hospital discharge register) and various study materials. The incidence completeness estimates are over 99%, and the follow-up of the patients is practically complete (Teppo et al. 1994, Dickman et al. 1999). For any analyses, it is also important that the completeness of registration does not vary by time, geographic area or population subgroup.

USES OF THE REGISTRY DATA

The Finnish Cancer Registry’s data are used to provide the Cancer Society of Finland and the authorities with the information basis needed for the actions in cancer policy. The scientists receive data for their projects on cancer causes and prevention that will ultimately benefit the society and the mankind. Clinical use of the database concerning decisions on individual patients is, according to law, however, not possible.

CONTENTS OF THE REGISTER DATA

The essential information in the cancer register is rather limited: On the person, the information includes the identity (personal identification number), municipality of residence and the date of death. On the tumour, the variables registered include the location, date of diagnosis, histological type and the address where to find more information. Very few clinically relevant variables in addition to tumour stage are registered.

CANCER PREVALENCE ESTIMATION

Because of the almost complete follow-up and the long-term existence of the cancer registry, not only the incidence but also the prevalence of cancer should be directly estimable (Möller et al. 2003). The prevalence, however, cannot be inferred based on the data of the cancer registry, as there are no data on the cure of the single registered cancer patients. The patients’ excess mortality compared to a comparable general population group with respect to gender, age, time period and possibly also region or social group can be estimated. In many cancers it is possible to estimate a time point after diagnosis after which the patients do not experience any excess mortality (Hakulinen et al. 1981). This is called the point of (statistical) cure.

The point of cure refers to the whole group of patients. However, the individual patients become cured before the statistical point of cure. The cancer register database, unfortunately, does not contain information on the basis of which the cure of an individual patient could be deduced. In clinical practice, this is not even possible for most patients, either, regardless of the clinical additional information at hand.
The excess mortality of the patients may be understood as a surrogate of the mortality specific to the primary cancer of the patients. The advantage of this concept is that it does not depend on the quality of the death certification. It also correctly accounts for the cancer-related deaths that are not necessarily attributed to the primary cancer, e.g., suicides and deaths due to the multiple tumours. On the other hand it may be difficult to find a proper general population group matching, e.g., for social class or smoking habits.

For most cancers, the excess mortality nevertheless disappears between five and ten years after diagnosis and its size starts to be relatively small already at five years. Thus, often cancer prevalence is defined as the proportion of living cancer patients in the population, with diagnosis during the latest five-year period. In Finland, on January 1, 2010, there were approximately 90,000 such persons in a population of 5.3 million giving a prevalence of about 17 persons per 1000 residents (Finnish Cancer Registry 2011). With a ten-year cutpoint, the number of prevalent cases would be 145,000 and the prevalence proportion 27/1000. Without any cutpoint the number is increased to 224,000 and the prevalence to 42/1000. The prevalence proportions vary very strongly by age and the overall prevalence figures are greatly influenced by the age distribution of the population. Note that the figures above are on cancer and not on patient prevalence. A living patient with more than one tumour is counted once per each tumour. Also the slightly smaller patient prevalence figures where each patient is included only once can be produced.

CANCER SURVIVAL AND MORTALITY

The excess mortality of the patients taken as the only mortality hazard of the patients defines a cumulative survival function called the relative survival (Hakulinen et al. 2011). This function gives an estimate for the patients’ survival when the effect of the competing risks of death has been eliminated, often called, the net survival of the patients (Pohar Perme et al. 2012). These net survival figures have the advantage that they, at least age-specifically, are comparable between different populations, as the eliminated mortality due to the competing risks of death is likely to vary between populations. The disadvantage is that the net survival exaggerates the real survival of the patients, sometimes called crude survival. For example, for male colon-cancer patients followed up in 2005-2007 in Finland, the five-year relative survival is 60% whereas the corresponding crude survival is 47% (Finnish Cancer Registry 2011). These figures differ markedly by age. For colon-cancer patients aged 75 years or more at diagnosis the five-year relative survival is 57% and the corresponding crude survival 34%.

The theory of competing risks of death can be used to split the complement of the crude survival between the two causes of death, patients’ cancer and the other causes of death, even when the actual causes of death of the patients are not available or reliable. In male colon cancer patients followed-up in 2005-2007, the five-year cause-specific death probabilities were 37% for colon cancer and 16% for the other causes of death. In the age category 75 years or more at diagnosis these probabilities were 39% and 27% for colon cancer and the other causes, respectively.

CANCER PREDICTIONS

It is important to be able to predict what to expect in the future. Accurate predictions are needed for planning the resources. They can also serve as baselines for evaluations of actions that have been made against cancer. An outcome lower than that predicted would
then count as a success of that action. Short-term predictions are very useful for estimating the incidence of cancer in the current year, since as a rule the latest published actual cancer statistics are based on two-three years older data. For example, the cancer incidence statistics for 2009 was published at the end of 2011 and reported 14731 new cases in males (Finnish Cancer Registry 2011). The same publication also included a prediction for 2012 with 15814 new cases (95% prediction interval from 15311 to 16315).

The unknown causes of death or the unknown distributions of the few known risk or causal factors in the population are a problem for making statistical models for predicting the occurrence of various cancers (Hakulinen 1996). Varying latency times from cause to effect make an additional problem. With the exception of smoking-dependent lung cancer and cancers with early detection or screening programs (breast, colon, cervix uteri, prostate), the best strategy is to make trend-based predictions (Dyba and Hakulinen 2008). Smooth changes in the effects of the unknown aetiological factors are then incorporated in the model parameters (age, period and cohort). The simple extrapolation models are not complicated by the well-known non-identifiability problem (birth year + age= year of observation) (Clayton and Schifflers 1987) and often work up to ten years into the future. Trends on which the predictions are based may also, however, be caused by (smooth) changes in, e.g., definitions, diagnostic criteria and facilities, and it is then questionable whether these trends can be extrapolated for predictions (Hakulinen 1996). It should also be noted that reporting to the cancer registry may be delayed and the last year of observation may still be incomplete for prediction purposes (Teppo et al. 1994). Accuracy is also needed in stating what has been predicted, age-standardised incidence or numbers of new cases. For example, the latest male cancer incidence prediction up to 15 years in to the future in Finland (Finnish Cancer Registry 2009) is a 7% decrease (age-standardised incidence), which is translated into a 29% increase in the predicted annual new cancer cases, because of the strong aging of the population.

The incidence predictions can be combined with forecasts on general mortality and predictions for patients’ excess mortality in order to give predictions for cancer mortality and prevalence (Hakulinen et al. 1989, Engeland et al. 1995, Verdecchia et al. 2002). These can be refined to include regions and subpopulations and used also as a basis for estimating future cancer costs (Mäklin and Rissanen 2006).

FINAL COMMENT

A full utilisation of the Cancer Registry’s database requires a good quality of data, appropriate scientific approach and rigour in the dissemination of the results. Keeping the database, statistics and research activities combined improves the quality of the material and makes it possible to use the material for decision-making and scientific goals.
REFERENCES


