A modification to the flow method to estimate completeness in cancer registries with delayed registration

Fabio Montanaro[1], David Robinson[2], Andrea Bordoni[3], Jean-Michel Lutz[4]

[1]Liguria Mesothelioma Registry, Department Epidemiology and Prevention, National Cancer Research Institute, Largo R. Benzi 10, 16132 Genova, Italy
[2]Thames Cancer Registry, King's College, 1st Floor Capital House, 42 Weston Street, London SE1 3QD, UK
[3]Ticino Cancer Registry, Institute of Pathology, Via in Selva 24, CH-6600 Locarno, Switzerland
[4]Association Suisse de Registres des Tumeurs, Registre genevois des tumeurs, 55 Bd de la Cluse, CH-1205 Geneva, Switzerland
Address correspondence to Fabio Montanaro, E-mail: fabio.montanaro@tele2.it

ABSTRACT

Objective For registries that routinely delay case registration, the flow method does not accurately describe their completeness over time. A modification to allow for such a delay is proposed and tested.

Methods Using original data set from Ticino Cancer Registry, Switzerland, a new data set (the delayed data set) was created by adding two years to the date of registration for each case in the original data set, to emulate the situation in a registry where registration is delayed by two years. Both the original and modified methods were then applied to both the original and the delayed data set.

Results When applied to the delayed data, the original method produced estimates of completeness of 32 and 43% at one and two years after diagnosis. When the modified method was applied to the delayed data, the completeness at one and two years was correctly estimated at 0%. After the initial two-year time lag, completeness was consistently estimated by both methods. When applied to the original data, the modified method produced the same results as the original method.

Conclusions The proposed modification allows the method to be applied even when registration is delayed long after diagnosis—thus extending the range of registries for which the flow method can be used.

Keywords cancer registration, completeness, quality control

Introduction

The estimation of the completeness of registration is an important criterion for assessing the quality of the data released by cancer registries. Several methods are in use, according to specific organization and local facilities.[1] Besides classical methods, the so-called flow method was introduced by the Thames Cancer Registry in 1995[2] and is now in use by several cancer registries in Europe.[3] A great benefit of this method is the description it provides of how completeness increases with time since diagnosis. However, the program currently available to analyse completeness with this method do not appropriately describe the data when cancer registration routinely begins with a delay (e.g. starting 1 or more years after the end of the year of incidence), as in most Italian registries.

We describe a modification to the programs, which allows them to be used independently of the actual delay in starting registration and, therefore, in a wider setting of cancer registries.

Materials and methods

The flow method

The flow method is based on the concept that registration is a time-dependent event observed after diagnosis and follows a probabilistic approach. In particular, it combines the following three time-dependent probabilities, obtained as described: $s(t_i)$ is the probability that a patient diagnosed with cancer is still alive at time t_i after diagnosis, obtained from the survival distribution; $m(t_i)$ is the probability that the death certificate of a patient who dies in the interval (t_i, t_{i+1}) includes a mention of cancer, obtained from cancer registrations of deceased patients; and $u(t_i)$ is the probability that a patient surviving until time t_i after diagnosis is still unregistered,

Fabio Montanaro, Consultant Epidemiologist

David Robinson, Honorary Senior Lecturer

Andrea Bordoni, Head of Ticino Cancer Registry

Jean-Michel Lutz, Epidemiologist

obtained using standard survival methods by treating registration before death as the event and censoring at death.

From these probabilities, it is possible to compute the probability of a patient's being registered, missing or lost. Complete details can be found in the original paper.[2]

Two Stata[4] ado-files have been made available by the authors: (i) comp.ado, which prints out estimates of completeness at each year after diagnosis (from 1 to 5) and related graphs; and (ii) complims.ado, which produces the same output as the previous program but, in addition, allows the user to specify the time points at which completeness is to be estimated and also provides 95% confidence limits for these estimates, using bootstrapping methodology.[5] Two data files are required as input to the completeness program: first a file containing all cancer cases on the registry's database diagnosed in a given year or period (the incidence file) and a second one containing all cases included in the registry's database who died during a given year or period, regardless of diagnosis year (the death file).

Proposed modification

The modification described here implies an adjustment in the calculation of missing cases in the period when no cases can have been registered.

The proportion of missing cases M at time t_i after diagnosis is given by:

$$M(t_i) = \text{prob(surviving and still being unregistered at time } t_i) = s(t_i) \cdot u(t_i)$$

where $s(t_i)$ is the probability that a cancer patient is still surviving at time t_i after diagnosis and $u(t_i)$ the probability that a patient surviving until time t_i after diagnosis is still unregistered. The proportion of missing cases $[M(t_i)]$ calculated in such a way can never be 1, as the survival is correctly estimated since the date of diagnosis, regardless of whether death occurs before or after the first possible registration date.

This limitation can be easily overcome by replacing $M(t_i) = 1$ for each time point where $u(t_i) = 1$.

This enables the program to estimate the time delay between diagnosis and registration, without requiring prespecification on the part of the user.

Test databases

Data collected by Ticino Cancer Registry (TCR), Switzerland, have been used to test the efficacy of the modification described above. TCR began its activity in 1996, collecting all new tumour cases occurring in the population of Canton Ticino (http://www.ti.ch/cancer/).

First, the completeness of cancer registration in Ticino was estimated from the original database (including both sexes and all cancer sites), using an incidence file of cases diagnosed in

1996 ($n = 1489$) and a death file of cases who died in 2000 ($n = 805$), with follow-up until 31 December 2000.

Next, a new database (the delayed database) was created in which the date of registration was modified by adding 2 years to the original registration date. This was done to emulate the situation in a cancer registry in which registration is delayed by 2 years. Follow-up was also extended by 2 years to 31 December 2002.

Both the original program and the modified ones were then applied to both the original and the delayed databases.

Results

Table 1 and Fig. 1 show the results obtained by both methods (original and modified method) using both data sets (original and delayed data).

Using the original method, the original TCR data showed a completeness of registration of 94% at 1 year and 97% at 5 years after cancer diagnosis (Fig. 1a). However, when the delayed data were analysed, the unsuitability of the original method was clear. As we had delayed the registration date by 2 years, no registered cases would be expected within the first 2 years, but 32 and 43% of the cases appeared to have been registered after 1 and 2 years, respectively (Fig. 1b). These estimates were clearly incorrect. On the contrary, when the modified method was applied to the delayed data, the completeness at both 1 and 2 years was correctly estimated at 0% (Fig. 1d). After this initial lag time, the completeness was consistently estimated by both methods.

When the modified method was applied to the original TCR data, we obtained exactly the same results as returned by the original flow method (Fig. 1c).

Discussion

The original flow method correctly described the change in completeness over time in the original TCR data. On the contrary, it was clear that it was not able to describe the

Table 1 Completeness (%) estimated by original and modified method using original data and data with registration date delayed by 2 years

Completeness (%)	Original method		Modified method	
	Original data	Delayed data	Original data	Delayed data
At 1 year	94.02	32.15	94.02	00.00
At 2 years	95.77	42.76	95.77	00.00
At 3 years	96.65	89.81	96.65	89.81
At 4 years	96.81	90.27	96.81	90.27
At 5 years	96.91	90.52	96.91	90.52
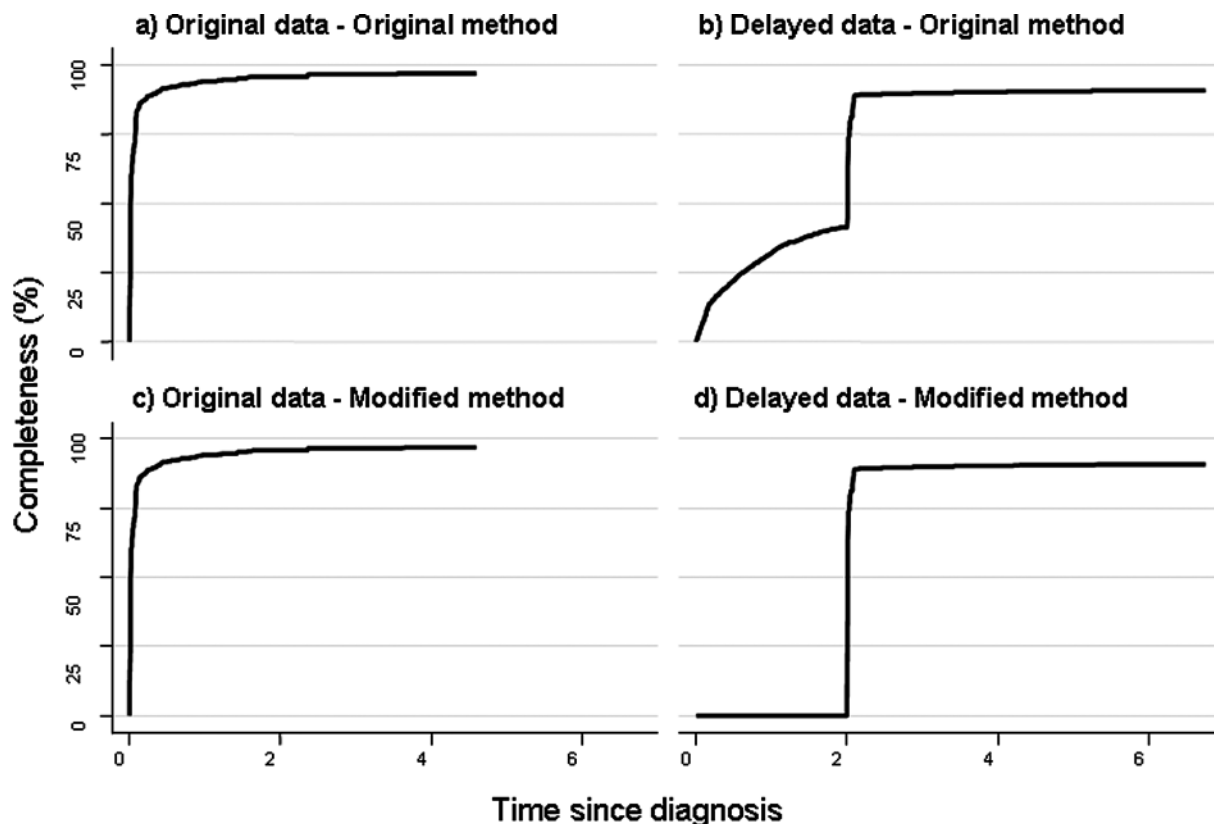
Fig. 1 Completeness estimated by original (**a** and **b**) and modified (**c** and **d**) method using original data (**a**–**c**) and data with registration date artificially delayed by two years (**b** and **d**).

situation where an initial delay in registration was observed. In fact, the original flow method gave a non-zero estimate of completeness in the period when no registration could have occurred at all. The modification introduced into the program allowed us to assign a non-zero probability of registration only after the first patient was actually registered. The same estimates of completeness were obtained by both methods >2 years after diagnosis, when the effect of the delayed start of registration no longer has any impact.

The survival experience of cases collected by registries that usually begin registration some time after diagnosis does not influence completeness estimates during this initial period, but its influence becomes important after registration begins. The advantage of the modified software is that it can be used to describe the completeness of both types of cancer registry (with real-time registration and delayed registration, respectively).

Conclusions

The modification introduced into the flow method program allows us to apply this method even when computerized registration is delayed long after cancer diagnosis.

With this modification, the flow method is able to describe a wider set of cancer registry data than the original program could do. It is important to stress that the results obtained with the modified method are consistent with those obtained from the original one when registration starts with no (or very short) delay after diagnosis. The use of the modified flow method program can be recommended for both real-time and delayed case registration.

Acknowledgements

References

1 Parkin DM, Chen VW, Ferlay J *et al.* Comparability and quality control in cancer registration. IARC Technical Report 19. Lyon: International Agency for Research on Cancer (IARC), 2004.

2 Bullard J, Coleman MP, Robinson D *et al.* Completeness of cancer registration: a new method for routine use. *Br J Cancer* 2000;**82**:1111–6.

3 Montanaro F, Ess S, De Weck D *et al*. Evaluating completeness of cancer registration in Switzerland. XXX Reunion du Groupe de coordination pour l'epidemiologie et l'enregistrement du cancer dans le pays de langue latine (GRELL). Lisboa, 2005. http://www.grellnet.org/2005/diapos/o19montanaro.pdf (3 July 2006, date last accessed).

4 StataCorp. *Stata Statistical Software: Release 8.0*. College Station: Stata Corporation, 2003.

5 Silcocks PB, Robinson D. Completeness of ascertainment by cancer registries: putting bounds on the number of missing cases. *J Public Health* 2004;**26**:161–7.