

## COMMENTARY

# Population Based Studies of Cancer Survival: Scope for the Developing Countries

Murali Dhar\*, Shivani Rao, R Vijaysimha

### Abstract

Survival refers to the life of a person after diagnosis of disease, and survival studies have the objective of evaluating the overall performance of a group of patients in terms of quality and quantity of life after the diagnosis or treatment. Potentially there are two approaches for the study of population-based survival; direct (classical) and indirect. The direct approach refers to defining a cohort of patients and collecting follow-up information, whereas the indirect approach uses current data on incidence and mortality for estimating various segments of life. In general, there are numerous difficulties in the conduct of population-based survival studies by the classical method, especially in the set-up of developing countries. These include time and finance required for the conduct of the study, the problem of loss to follow-up and also the time gap between the year of diagnosis of patients and the availability of results on their survival. In fact the problem of time gap is recognized even in the developed world. There have been many studies highlighting these problems and suggesting potential solutions. Generally they have focused on three directions: viz, improving the address information and thereby reducing the loss to follow-up; development of methodologies to deal with the losses to follow-up and indirect studies of cancer survival, thereby obviating the follow-up process. This commentary covers the potential approaches of population-based survival studies, classical survival studies, problems in the same, and methods adopted for their solution. A summary of the conceptual and methodological developments on these concepts, highlighting the scope for the developing countries, is also presented.

**Keywords:** Cancer - survival study - classical approach - indirect approach - India - developing countries

*Asian Pacific J Cancer Prev*, **11**, 831-838

### Introduction

Knowledge of survival is essential in the community level management of a disease. It's knowledge over a period of time helps in monitoring and improving the levels of prognostic factors in the population. In addition, survival duration also helps in deriving various indices of burden of disease, like, disability adjusted life years, healthy life expectancy, etc. There are many potential approaches in the study of survival of a disease possessing their own merits and limitations. The selection of a particular approach depends on many factors, like, specific objectives of the study, time available for conducting the study, finance available, availability of analytic tools, etc. In general, there are numerous difficulties in the conduct of survival studies by the direct method especially in the set up of developing countries. These include, time and finance required for the conduct of the study, the problem of loss to follow-up and also the time gap between the year of diagnosis of patients and the availability of results on their survival. In fact the problem of time gap has been realized even in the developed world. Due to these difficulties, only sporadic survival studies have taken place in the developing countries. In India, there are survival

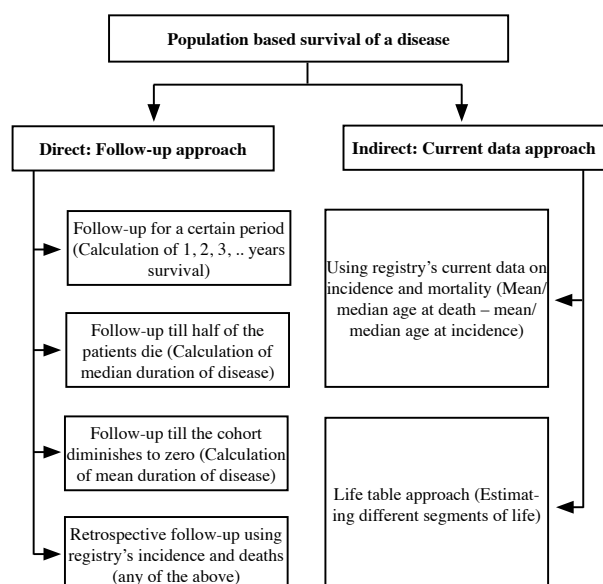
studies associating the level of prognostic factors with the survival. However, there is no study comparing the level of prognostic factors over a period of time in correspondence with the survival. For example, there are studies reporting education and stage of the disease as the independent predictors of survival. But there is no study correlating the trends in the levels of education and stage of disease with the survival rates. As far as survival duration is concerned, there is no visible attempt of direct estimation in the set up of developing countries probably due to lack of follow-up information. There are, however, few attempts employing indirect techniques.

This study presents a detailed picture of potential approaches of population based survival studies, classical survival studies, problems in the same, and methods adopted to solve these problems. A summary of the conceptual and methodological developments on these concepts has also been presented.

### Basic Approaches in the Study of Cancer Survival

Figure 1 presents potential approaches of a population based study of survival of a disease. Broadly, there are

Department of Community Medicine, JSS Medical College (A constituent college of JSS University), Mysore, India. \* For correspondence : murali\_dhar\_1966@yahoo.com, mmdhhaarr@yahoo.co.in



**Figure 1. Diagrammatic Presentation of Potential Approaches of a Population Based Study of Survival from a Disease.**

two approaches: direct (follow-up) approach and indirect (current data) approach.

In direct approach, which may also be called as classical approach, one needs to define a cohort of patients and follow them prospectively or retrospectively for certain duration. In a prospective follow-up study, duration of follow-up depends on many factors like, finance, time, disease history, etc. If we are interested in survival proportion at a certain interval of time, we have to follow the cohort for that interval of time. If we are interested in calculating median duration of disease, we have to follow the cohort till half of the patients die. If we are interested in calculating mean duration of disease, we have to follow the cohort till the cohort diminishes to zero. On the other hand in a retrospective survival, we can make use of the registry data over a period of time to match the cases of incidence in the past with the deaths. Once the matching is done, we can calculate any of the three measures of survival discussed above.

Indirect approaches have been developed recently and are suitable especially for the setup of developing countries. In these approaches, we make use of current data (i.e., data of a particular year) on incidence and mortality, rather than follow-up data. The validity of the results of this approach depends on the stability of the incidence and mortality rates in the particular population under consideration. For the purpose of this approach, we calculate hypothetical value of the duration of disease under the assumption of the stability of incidence and mortality rates. This approach is similar to the approach of studying the loss in the presence or the gain in the absence of a particular cause of death. In studying the loss/gain, we are interested in age at death in two circumstances; one, when all the causes of death are in force and two, in the assumed absence of a particular cause of death. The concept of loss or gain from a particular cause of death has already been reviewed in detail elsewhere (Dhar et al, 2008b)

## Classical Approach of Survival Study

The traditional definition of survival refers to the life of a person after the diagnosis of the disease under study. It should ideally be measured by the number of years lived by a person after the diagnosis of the disease. To arrive at this type of measure however, one has to define a cohort of patients diagnosed with the disease, follow them till each and every one of the cohort dies and then calculate the average survival in terms of the average number of years of life lived by the cohort. Such an approach of survival study seems difficult because such a study requires long period of time to complete. Moreover, average survival thus obtained has its own demerit of being influenced by extreme values (outliers). These may be the reasons there is no visible survival study dealing with such an approach. Alternatively, survival studies calculate probabilities of the subjects surviving a particular duration after diagnosis; 1-year, 2-year, 5-year, etc. This approach consists of defining a cohort of patients, following them till a reasonable cut-off date and then applying a suitable method for the estimation of survival rates.

The collection of follow-up information for the survival studies can be active or passive. Active collection means the registry personnel visiting the houses of subjects under study and collecting vital status information. In case of many population-based cancer registries, collaborating hospitals or the concerned departments conduct annual follow-up surveys to collect the follow-up data and communicate the same to the registries (Hanai and Fujimoto, 1985; Parkin and Hakulinen, 1991; Esteve et al., 1994). Method of passive collection means the subjects or their representatives themselves coming for the follow-up of the disease and facilitating their vital status being recorded. This method requires the notification of the outcome under consideration by the vital registration system to the registries being mandatory. Any passive method helps in reduction of the cost to a substantial extent. However, if the notification through the vital system is not accurate, the estimates of survival rates based on passive follow-up data may be biased (Hanai and Fujimoto, 1985; Parkin and Hakulinen, 1991; Esteve et al., 1994).

## Methods for Estimation of Survival Rates

There are mainly three methods of estimating a survival rate; direct method, actuarial method and Kaplan-Meier (K-M) method. Direct method has been confined to literature due to the technical limitations attached to it in spite of being the simplest one in calculation and understanding. Actuarial method was in maximum use till few decades ago. Once the boom in the availability of computers eased the computational processes, the K-M method took the place as number one. Further, the idea of studying the impact of prognostic factors on the survival gave the rise to Cox model. Let us present a brief discussion on each of these methods individually. For a detailed picture of these methods, one may refer respective references.

This is the simplest method for estimating survival from follow-up data. Under this method, survival at the end of a fixed time period is calculated by dividing the number surviving at the end of the period by the number at the beginning of the period (Berkson and Gauge, 1950). This method uses the information on only those subjects who are followed for whole of the fixed time period under consideration. Censored observations are not considered at all. This may be the reason this method is termed as direct method (Mould, 1976). The main disadvantage of this method is that it does not make use of the censored observations. Survival calculation takes into account only those subjects who have been followed up from beginning till end of the specified time interval. Thus, information on the censored observations becomes redundant when using direct method. This is the reason survival for different time interval can not be based on same subjects, leading to another disadvantage of this method and that is, the survival may or may not decrease with time. The only advantage of this method is the simplicity in its calculation and understanding. The calculations by this method can easily be performed without any computer.

#### *The actuarial method*

Cutler and Ederer (1958) devised this method overcoming the disadvantages with direct method discussed above. This method makes use of all the follow-up information collected including those lost to follow-up and those censored on the date of termination of study. It gets the name from the fact that it is based on the principles and techniques of life tables, a tool devised and used by actuaries and demographers. Unlike direct method, this method involves few steps in estimation of survival over a time period. First step is to divide whole observation period in time intervals. Length of the intervals depends on the distribution of deaths over the observation period. Therefore, although not necessarily but generally time intervals are of equal length (1-year). Most of the studies have the observation period of 5 years divided in the intervals of a year thereby giving 1-year, 2-year, 3-year, 4-year and 5-year survivals. Once the time intervals are ascertained, next step is to calculate conditional probabilities of survival in the intervals. That is, the probability of a subject surviving and reaching the end of the interval given that he/she has reached the beginning of the interval. The numbers at the beginning of the intervals are adjusted by adding half of the subjects who were lost to follow-up or censored in that interval. The last step is to calculate the survival probabilities for the intervals by multiplying all conditional probabilities from first till that interval. In addition to overcoming the disadvantages with the direct method, this method also possesses the advantage of providing the survival pattern, i.e., the manner in which the subjects at the starting of the study diminish during the observation period. The drawback with this method is that in case of censored subjects, it does not use the follow-up information that falls in the censoring interval. This is because, this method assumes that censoring is evenly distributed within a time interval, thereby treating all the subjects censored in a particular interval to have been censored at exactly the

middle of that interval.

#### *The Kaplan-Meier method (Kaplan and Meier, 1958)*

This method is based on the same principles as the actuarial method. Calculation procedures are also similar. The main difference lies in the fixation of time intervals. Unlike actuarial method, the time intervals are not set a priori; rather, these are decided by the occurrence of the events during the observation period. The time gaps between two successive events become the time intervals. Thus, the number of time intervals and the number of calculations of conditional probabilities are as many as the number of distinctively occurring events. This feature makes the calculations relatively complex and it is considered to be specially suited for a small number of subjects. This may be the reason for relatively rare use of this method till a few decades ago. But the advancement in the computer technology and better accessibility of computers has left the issue of computational problems redundant. This is why, as of late, this method is in greater use.

#### *Corrected and relative survival rates*

The methods discussed above deal with the observed survival rate considering all the deaths irrespective of the cause of death. The reference point for the comparison of observed survival rate (in percentage) is 100. The deviation of observed survival from 100 quantifies the risk in the population under study. As the estimation of observed survival is based on all the deaths, this deviation has two components, the risk of death from the cause under study and the risk from other causes. The ideal way of removing the effect of other causes of death is to ignore the deaths from other causes and consider them as the censored observations. This is known as corrected survival rate. The main problem in calculating corrected survival rate is the non-availability of cause of death information. Alternate solution is relative survival rate, which is calculated by taking the ratio of observed survival rate of the patients in the study to the survival of the people with similar demographic characteristics in the general population (Ederer et al, 1961). Survival in general population is derived from life tables constructed employing the deaths from all causes of death. This is done under the assumption that the deaths from the considered cause are very small compared to deaths from all causes of death.

#### *Standardization of survival rates*

Age is almost a universal predictor of an outcome in epidemiological studies. This is true in case of survival also; survival from most (if not all) of the diseases tends to decrease with age. Thus when we compare the survival rate of two populations, the difference in the two, may have two components; one, the difference in the risk of death and two, the difference in the age structure of the patients included in the study. To eliminate the effect of age structure, one may opt for survival rates for smaller age intervals on the lines of age specific rates. However, an administrator or policy planner is generally interested in a single measure of comparison. To come out with

a single summary measure, direct standardization of relative survival rates has been advocated (Parkin and Hakulinen, 1991). The standardization is done on the lines of standardization of incidence or mortality rates; by using age specific survival rates in place of age specific incidence or mortality rates and standard cancer patient population in place of world standard population. To arrive at standard cancer patient population, Berrino et al (1995) in EUROCARE study combined the data from all the registries within each category of tumour site. On the other hand, Black and Bashir (1998) constructed a set of abstract world standard cancer patient population to facilitate comparisons among the survivals reported from the developed and developing countries.

#### *Other methods*

There are a few additional methods dealing with refinement, extension or special cases of above methodologies. We restrict here with just a mention of the same. Their detailed methodologies are beyond the scope of this study and are available in respective references. One can calculate the confidence interval of survival rates to assess the impact of sampling error (Greenwood, 1926; Rothman, 1978; Anderson et al., 1982). There are techniques developed to estimate mean survival time from incomplete observations (Gross and Clark, 1975; Hakama and Hakulinen, 1977). Median survival time can be calculated if at least half of the subjects have met the event within the observation period (Gross and Clark, 1975). Median survival time may also be approximated from 1, 3 and 5-years survival rates, normally reported by population based survival studies (Dhar et al., 2006). If cause of each death during the observation period is available, we can calculate cause specific survival (Chiang, 1968).

#### *Methods for comparing survival rates*

There are many tests in literature, developed to compare the difference between more than one survival curves (Mantel and Haenszel, 1959; Gehan, 1965; Breslow, 1970; Peto, 1972; Peto and Peto, 1972; Peto and Pike, 1973; Breslow, 1979), the most prominent one of those being the log rank test (Peto et al., 1977). Stratification of the subjects with inter-related prognostic factors and then employing rank tests is another way of comparing survival rates. The survival rates for different levels of a factor can be assessed adjusting for other prognostic factors by employing a stratified rank test (Mantel, 1966). The impact of a particular factor on survival adjusted for other confounding factors can be studied by applying Cox proportional hazard's model (Cox, 1972).

### **Drawbacks with the Classical Method of Survival Studies**

There are numerous logistic and technical problems in the conduct of prospective or retrospective survival studies based on follow-up data in the setup of developing countries. Important ones are financial requirement, time needed and loss to follow-up. The first two, although the main reason for limited number of survival studies in

developing countries, have escaped enough mention in the literature. Another problem, which has been realized even in the setup of developed world, is the time gap in the reporting of survival.

#### *Technical*

The disease management efforts can be classified broadly into two groups. Prevention efforts aimed at reducing the incidence rate and there by increasing the average survival before contact with the disease and the control efforts aimed at improving the quality and quantity of survival after the diagnosis of the disease. Classical method of survival study reflects the effects of control measures taken in the population under study, i.e., the improvement in the survival after diagnosis of the disease. However, it does not deal with the effects of prevention efforts adding to the disease free life. In fact, additions in disease free survival are more important than additions in with-disease or disease-free survival after diagnosis of the disease, because, years of life gained by prevention are more likely to be fully productive than years of life gained through medical care and therapy (Arca et al., 1988).

Another drawback with the conventional method of survival study is that it deals with the survival experience of the subjects for a limited period, generally 5-years. Thus if the 5-year survival rate in a particular study is 50%, it means study dealt with complete survival experience of only half of the patients. About remaining half of the patients, the study can highlight only that they lived at least for 5 years. The conventional method is silent on what happens to the half of the subjects after 5 years of their survival.

#### *Finance and time*

Mathew (1996) rightly stated, "Lengthy periods of observation may be required until an event occurs, and the maintenance of surveillance on patients within the study group can be extremely costly and time-consuming". Thus, financial requirements and time are important considerations in a survival study. These are more so in a population based survival study and even more so in case of a prospective study. In a survival study, the cohort needs to be followed till a sufficient number of events occur in order to provide scientifically meaningful and reliable survival probabilities. In case of cancer, one has to follow at least for 5 years to arrive at 5-year survival rates, commonly reported. Even in a retrospective survival study, quite a long time (may be a year or more) is required to collect data for whole observation period, which is again minimum 5 years in case of cancer.

As far as financial requirements are concerned, there is no visible study dealing with the evaluation of cost of a survival study. This may be because cost of a survival study has many components from planning, employing trained personnel to collect data, acquiring suitable computer software for analyzing and reporting. The cost of a population based survival study also depends on many factors, such as, the number of subjects, scattering of the subjects in the study population, whether the study is prospective or retrospective, etc. Thus it is difficult to work out a uniform cost pattern for the

survival studies. However, it is unarguably a substantial amount, affordability of which is difficult in the set up of developing countries. This is why only a few populations in India have been reported with few survival studies, that too only in the recent past. A number of these were possible probably only due to financial aid from abroad (Gajalakshmi et al., 1997).

#### *Loss to follow-up*

The main goal in any survival study is to obtain complete follow-up information. But at the same time, the problem of loss to follow-up (LFU) is inherent in any survival study. There is no problem if the losses occur randomly with respect to the factors related with the event because then the losses will be in compliance with the assumptions involved in estimation of survival rates. The effect of loss to follow-up on the survival rates is two-dimensional; the extent of non-randomness of LFUs and the quantum of LFUs, and both of these, i.e., LFUs being related and LFUs being substantial in number are well established. The effect of LFUs on the survival rates is mostly in opposite direction in the developed and developing countries. In the developed countries, as the death registration is almost complete, dead patients are more likely to be traced and included in the study. Thus LFUs cause under-estimation of survival rates in developed countries (Enstrom and Austin, 1977; Austin, 1983). On the other hand in the setup of developing countries, where death registration is incomplete, death plays an important role in a subject being lost from the follow-up. Thus, a dead subject is less likely to be traced and get included in the study and therefore LFUs cause over-estimation of survival rates in the set up of developing countries. As far as the quantum of LFUs in the developing countries is concerned, there are reports of up to half of the patients being lost to follow-up in the survival studies. Fifty percent of patients in a survival study on esophageal cancer were lost within first year of follow-up (Desai et al., 1969). In a study on head and neck cancer patients, fifty percent of patients were lost to follow-up within first three years of follow-up (End result report, 1990). Ganesh (1995) reported 24 percent lost to follow-up within three-year period of the study on breast cancer patients. In a survival study on ovarian cancer, Mathew (1996) reported more than 40% of LFUs within first year and two third in five years. Many researchers have found the bias in the estimates of survival due to LFU (Berkson and Gage, 1950; Cutler and Ederer, 1958; Drolette, 1975; Enstrom and Austin, 1977; Austin, 1983; Tallis et al., 1988; Ganesh, 1995; Mathew, 1996). These studies have shown that survival rates are grossly misreported if not adjusted for the effect of LFUs. In a study of survival of ovarian cancer patients registered at Regional Cancer Centre, Thiruvananthapuram, 5-year survival rate declined from 75 to 43 percent when adjusted for LFUs, a decline in the survival rate of over 40 percent (Mathew, 1996).

#### *Difficulties in obtaining follow-up information*

India is a vast country with socio-cultural diversity. The absolute burden of cancer has been increasing due to

increase in overall population and in relative proportion of elderly population as a result of control of communicable diseases. Due to this, data on cancer in India have been limited and that on survival are rare. There are only a few reports on cancer survival in India. There are a number of problems in obtaining follow-up information in most part of the developing world including India. Only a few of the hospitals have an organized and regulated system for follow-up. Further the need for follow-up is not appreciated by most of the patients in our setup. Furthermore, patients have their own logistic, economic and socio-demographic constraints in responding.

The place of residence plays an important role in access to the health resources. A high proportion (72%) of Indian population lives in rural areas (Health Information India, 2005). Quite often, patients from rural area reach the hospital for initial treatment and do not turn back for further medical care or follow-up. The inconvenience and cost involved in travel are some of the reasons for such non-responses.

An important reason for lost to follow-up is incomplete treatment. It is widely held in India that one third of all cancers are preventable and another one-third curable if detected early. The result of any treatment depends on the completion of the prescribed course of treatment. The proportion of patients dropping out with incomplete treatment is quite considerable in the country especially in case of cancer. This may be due to one or more of many possible reasons, like, side effects of treatment, high cost involved, relief from symptoms and ignorance about the need for completing treatment, etc.

Accessibility of hospital by a cancer patient is another problem in our country. Although we have well-organized health infrastructure in the country, the facilities for the diagnosis and treatment of cancer are not sufficient. Moreover the facilities available are concentrated in urban centers leading to patients from rural areas having a long distance to travel. This along with their ignorance about the importance of follow-up almost inhibits them from coming for follow-up.

#### *Time gap*

The drawbacks with the traditional method of survival study discussed above are by and large specific to only developing countries with scarce financial resources and poor (or even no) health information system. The problem of time-gap however, is an issue in case of developed world also. This problem refers to the duration between the diagnosis of the patient and reporting of their survival and results in a substantial delay in the reporting of changes in prognosis over time. It was in 1996, when Brenner and Gefellar highlighted this problem and suggested a new method to overcome the same.

## **Methods for Solving the Problem**

Potential solutions to the problems with the classical survival studies can be classified on the lines of the problems; cost and duration, loss to follow-up and time-gap.

#### *Murali Dhar et al* *Practical approaches to reduce the cost and duration*

The cost and the duration of a survival study are mainly the result of the collection of primary data on follow-up, prospective or retrospective. Thus avoiding the collection of follow-up information can substantively reduce the cost and duration of a survival study. This is not possible in the conventional method of study of survival because complete follow-up information is actually a pre-requisite of the same. In order to avoid collection of follow up information, one may explore the possibility of utilization of secondary data, because cost effective studies are achieved generally by using secondary data applying indirect techniques. As per our knowledge, there is no visible attempt so far in this direction barring few studies by Dhar et al., (2006; 2008a) recently.

#### *Approaches to solve the problem of loss to follow-up*

There are three stage efforts aimed at solving the problem of loss to follow-up; one, obtaining complete follow-up information given that address given by the patients are correct, two, obtaining correct address from the patients for further follow-up in the future and three, technical developments to deal with the bias due to the left over LFUs.

#### *Improving follow-up information*

Given that addresses recorded are correct, the theoretically easiest way to improve the follow-up information is to visit the patient's houses and collect the information on their vital status. But the ease does not come without substantial additional costs; visit to each patient's house increases the cost of the study substantially. Therefore house visit may not be an optimum option in improving the follow-up information. Considering the cost effectiveness of this strategy, it may not be possible to implement this system for the improvement of follow-up in many parts of the country (Mathew, 1996). Alternatively, several researchers made postal attempts to improve the information on loss to follow-up (Shanta and Gajalakshmi, 1989; Varghese et al, 1991; Sankaranarayana et al., 1995).

Shanta and Gajalakshmi (1989) collected up to seven addresses from the patient's attendant at the time of registration. If the patient did not turn up for follow-up on due date, reply paid post card was sent to patient's permanent address. If there was no response from the permanent address, reply paid post cards were sent to the remaining addresses successively at an interval of one month till a reply was received. The study was reported to be highly successful in improving the follow-up information.

Varghese et al (1991) sent reply paid post cards with instructions written in local language. They sent post cards to the house address of the patient and waited for 4 to 6 weeks and in case of no reply, cards were sent to alternate addresses. The follow-up information in that study improved from 33 to 68%.

Sankaranarayanan et al (1995) also used postal inquiries to improve follow-up information in the study of survival of cervical cancer patients treated at Regional Cancer Centre, Thiruvananthapuram.

#### *Improving the address information*

Accuracy of address is an important issue in initially obtaining or subsequently improving the follow-up information. It was the concern of the accuracy of addresses that must have prompted Shanta and Gajalakshmi (1989) to collect up to seven addresses from patient's attendants. There are studies in the past to test the accuracy of the addresses reported by the patients and the reasons for inaccuracy (Krishnaswami et al., 1979; Radhakrishna et al., 1980; Satagopan et al., 1983). Krishnaswami et al., (1979) looked into the accuracy of home addresses given by the patients of three tuberculosis clinics in Chennai, a metropolitan city with better civic system. The accuracy of the addresses was found to be very poor, with only 70 to 80 percent of the letters posted reaching the patients. Supplementation of the efforts by appointing a well-motivated and experienced staff at the registration desk increased the accuracy of the address by 10 to 20 percent. Radhakrishna et al (1980) investigated the factors influencing the accuracy of the addresses reported by the patients. The accuracy of the addresses was found to be substantially low for the illiterate patients and for those who did not have a permanent place of residence. The role of education may be visualized by the fact that in this study 98% of the cards, on which, address was recorded by a literate person, were returned by the patients.

#### *Technical developments to deal with incomplete follow-up*

Finally, when we end up with a substantial amount of losses to follow-up in spite of all the efforts discussed above, there are technical developments to deal with the same and come out with the adjusted survival (Tallis et al., 1993; Geanesh, 1995; Mathew, 1996). These developments are based on their own assumption about behaviour of risk of the event with respect to different level of various prognostic factors. Moreover, these adjustments generally necessitate the collection of data of more variables in addition to the variables normally needed for the study of survival.

Tallis et al (1993) proposed a procedure for adjusting for potential bias in LFUs that may affect the registries using passive method of follow-up for survival analysis. Ganesh (1995) suggested methods for adjusting survival rates for the effect of non-random LFUs by stratification of the subjects according to different levels of prognostic factors. Mathew (1996) also came out with a method for loss adjusted survival rate through a statistical model based probability approach.

#### *Indirect method*

In indirect approach, we make use of current data (i.e., data of a particular year) on incidence and mortality, rather than follow-up data. The validity of the results of this approach depends on the stability of the incidence and mortality rates in the population under study. Dhar et al., (2006) estimated duration of disease for selected sites of cancer using the data on cancer incidence and mortality from PBCR Mumbai. They estimated median duration of disease by subtracting median age at incidence from median age at death. The main methodological disadvantage was that while calculating median age at death, the cancer patients dying from other than cancer

were not considered. Under estimation due to this was reported to be 30 to 40 percent.

Later on, Dhar et al., (2008a) proposed an indirect methodology for the study of survival. It was reported to be suitable especially for the setup of developing countries, where financial resources are scarce and an effective and comprehensive follow-up system is lacking. The method is based on estimating various segments of life employing life table techniques and comparing the same over a period of time. The main strength of the proposed procedure is that, unlike traditional method, it does not require the follow-up information for a cohort of patients. Thus it does not involve the difficulties related to finance, time and loss to follow-up, which are inherent with the traditional method and have been discussed in detail in earlier sections.

#### Period analysis

The problem of time gap with the classical method was realized during mid nineties, when Brenner and Gefeller (1996) came out with the idea of period monitoring instead of usual practice of cohort monitoring. The conceptual idea was conceived from the usual practice of period life tables to estimate current life expectancy instead of cohort life tables providing redundant information on expectation of life. Subsequently this method has been used with varying terminology, like, 'period analysis' (Brenner and Gefeller, 1997) and 'model based period analysis' (Brenner and Hakulinen, 2006). However, ultimate objective has been same and that is to overcome or at least substantially reduce the limitations with the traditional method. Basically the period analysis approach uses information from the same patients (say, recruited between 1980-90), but the analysis is restricted to the numbers at risk and the events contributed by these patients during a more recent calendar period, say, 2000-2005, thereby discarding survival experience observed during earlier years. This resulted in more up-to-date estimates of cumulative survival on the cost of loss in precision compared with complete analysis.

#### Summary

Knowledge of survival is essential in the community level management of a disease. Broadly, there are two approaches of population-based study of survival from a disease, the direct (i.e., classical) approach and the indirect approach. In classical approach, survival refers to the life of a person after diagnosis of disease, and population based survival studies deal with the measurement of the same to evaluate overall performance of a group of patients in terms of quality and quantity of life after diagnosis/treatment. Classical approach ignores the improvement in survival before contact with the disease resulting from preventive measures or improvement in general health awareness. Besides, there are numerous difficulties in the conduct of a population based survival study in the context of developing countries including India. While planning a survival study, one has to consider the possibility of substantial amount of financial and other resources required including time. Subsequently, loss to follow-

up is a typical problem encountered in survival studies especially in developing countries. The losses to follow-up have been reported to be substantial causing biased estimates. There have been quite a few methodological researches into the solutions for the above problems with classical approach. However, none of the methodological developments has been able to acquire the status of established accepted method. Another problem with the classical approach, which has been realized even in the developed world, is the time gap between the dates of diagnosis and reporting of the survival. In fact the realization of this problem resulted in the development of concept of period analysis.

On the other hand in indirect approach, one may use the current data (i.e., data of a particular year) on incidence and mortality rather than follow-up data and estimate various segments of life. The validity of the results of this approach depends on the stability of the incidence and mortality rates in the population under consideration. The indirect method suggested by Dhar et al consists of constructing three life tables by applying various attrition factors: (a) risk of death from all causes; (b) risk of incidence and that of death from other causes; and (c) risk of death from other causes only. Various segments of life are then arrived at from three different expectations of life and suitable subtractions among them. Given the difficulties in conduct of classical survival studies, the indirect method may provide a useful tool for the study of cancer survival viable for the developing countries. Indirect method may also be useful in having a regular audit of the prognostic factors prevailing in a population. Further, it also has potentials to be utilized for the estimation of various indices of burden of disease, like, PYLL, DALY, etc.

In conclusion, the classical method is no doubt the ideal method for the study of survival. However, feasibility of the same depends upon many considerations, important ones being availability of enough finance and an adequate health information system. Both of these are, by and large, lacking in the set-up of developing countries. Therefore till a particular developing country comes upto the expected level in terms of financial ability and health information system, indirect methods may be used in the study of survival especially in case of dealing with large populations. Even otherwise, indirect method may be useful as a quick and fast method for a regular audit of the prognostic factors.

#### References

- Anderson JR, Berstein L, Pike MC (1982). Approximate confidence intervals for probabilities of survival and quantiles in life table analysis. *Biometrics*, **38**, 407-16.
- Arca M, di Orio F, Forastiere F, et al (1988). Years of potential life lost (YPLL) before age 65 in Italy. *Am J Public Hlth*, **78**, 1202-5.
- Austin CF (1983). Cancer registries: A tool in epidemiology, In: Review of cancer epidemiology, pp. 118-140. Ed. AM Lilienfeld, vol.2, Elsevier, New York.
- Berkson J, Gage RP (1950). Calculation of survival rates for cancer. *Proc Staff Meet Mayo Clin*, **25**, 270-86.

- Berrino F, Sant M, Verdecchia A, et al (1995). Survival of cancer patients in Europe: the EUROCARE study (IARC Scientific Publications No. 132), Lyon, France.
- Black RJ, Bashir SA (1998). World standard cancer patient populations: A resource for comparative analysis of survival data. *IARC Sci Publ*, **145**, 9-11.
- Brenner H, Gefeller O (1996). An alternative approach to monitoring cancer patient survival. *Cancer*, **78**, 2004-10.
- Brenner H, Gefeller O (1997). Deriving more up-to-date estimates of long-term patient survival. *J Clin Epidemiol*, **50**, 211-6.
- Brenner H, Hakulinen T (2006). Up-to-date and precise estimates of cancer patient survival: Model-based period analysis. *Am J Epidemiol*, **164**, 689-96.
- Breslow N (1970). A generalized Kruskal-Wallis test for comparing k samples subject to unequal patterns of censorship. *Biometrika*, **57**, 579-94.
- Breslow N (1979). Statistical methods for censored survival data. *Environ Hlth Perspect*, **32**, 181-92.
- Chiang CL (1968). Introduction to stochastic processes in Biostatistics. Wiley, New York, USA.
- Cox DR (1972). Regression model and life tables (with discussion). *J R Statist Soc B*, **34**, 187-220.
- Cutler SJ, Ederer F (1958). Maximum utilization of the life table method in analyzing survival. *J Chronic Dis*, **8**, 699-712.
- Desai PB, Borges EJ, Vohra VG, et al (1969). Cancer of the oesophagus in India. *Cancer*, **23**, 979-89.
- Dhar M, Lahiri S, Takiar R, et al (2008). An indirect study of cancer survival in the context of developing countries. *Asian Pac J Cancer Prev*, **9**, 479-86.
- Dhar M, Lahiri S, Vijayasimha R, et al (2008). Study of social loss or gain from a particular cause: Scope for the developing countries-a review with reference to cancer. *Asian Pac J Cancer Prev*, **9**, 797-804.
- Dhar M, Takiar R, Murthy NS (2006). A comparison of distribution of age at diagnosis with age at mortality of cancer cases - does it help in estimation of median duration of disease. *Cancer Registry Abstract*, **13**, 3-11.
- Drolette ME (1975). The effect of incomplete follow-up. *Biometrics*, **31**, 135-44.
- Ederer F, Axtell LM, Cutler SJ (1961). The relative survival rate: A statistical methodology. *Natl Cancer Inst Monogr*, **6**, 101-21.
- End Result Report (1990). End result report on head and neck cancer. Eds. PB Desai, DN Rao, RS Rao and PD Shroff. Tata Memorial Centre Scientific Publications, Tata Memorial Hospital, Mumbai, India.
- Enstrom JE, Austin FD (1977). Interpreting survival rates. *Sci*, **4**, 847-51.
- Esteve J, Benhamou E, Raymond L (1994). Statistical methods in cancer research Volume IV Descriptive epidemiology. *IARC Sci Publ*, **128**, 1-302.
- Gajalakshmi CK, Shanta V, Swaminathan R, et al (1997). A population-based survival study on female breast cancer in Madras, India. *Br J Cancer*, **75**, 771-5.
- Ganesh B (1995). Effect of loss to follow-up in estimating survival rates. Acta Universitatis Tampereensis. Ser A vol. 440, Tampere, Finland.
- Gehan EA (1965). A generalised Wilcoxon test for comparing arbitrarily singly censored samples. *Biometrika*, **52**, 203-23.
- Greenwood M (1926). A report on the natural duration of cancer reports in public health and medical subjects. *Ministry Hlth*, **33**, 1-26.
- Gross AJ, Clark VA (1975). Survival distributions: reliability applications in the biomedical sciences. John Wiley and Sons, New York.
- Hakama M, Hakulinen T (1977). Estimating the expectation of life in cancer survival studies with incomplete follow-up information. *J Chronic Dis*, **30**, 585-97.
- Hanai A, Fujimoto I (1985). Survival rate as an index in evaluating cancer control. *IARC Sci Publ*, **66**, 87-107.
- Health Information India (2005). Central Bureau of Health Intelligence, Directorate General of Health Services, Ministry of Health and Family Welfare, Government of India.
- Kaplan EL, Meier P (1958). Non-parametric estimation from incomplete observation. *J Am Stat Assoc*, **53**, 457-81.
- Krishnaswami KV, Satagopan MC, Somasundaram PR, et al (1979). An investigation of the accuracy of the home address given by patients in an urban community in South India. *Tubercle*, **60**, 1-11.
- Mantel N (1966). Evaluation of survival rates and two rank order statistics arising in its consideration. *Cancer Chemotherapy Report*, **50**, 163-70.
- Mantel N, Haenszel W (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst*, **22**, 719-48.
- Mathew A (1996). Removing bias in cancer survival estimates by active follow-up and information on determinants of loss to follow-up. Acta Universitatis Tampereensis. Ser A vol. 525, Tampere, Finland.
- Mould RF (1976). Calculation of survival rates by the life table and other methods. *Clin Radiol*, **27**, 33-8.
- Parkin DM, Hakulinen T (1991). Cancer registration: principles and methods. Analysis of survival. *IARC Sci Publ*, **95**, 159-76.
- Peto R (1972). Rank tests of maximal power against Lehmann, Type alternatives. *Biometrika*, **59**, 472-4.
- Peto R, Peto J (1972). Asymptotically efficient rank invariant test procedures with discussion. *J R Statist A*, **135**, 185-207.
- Peto R, Pike MC (1973). Conservation of approximation (O-E)<sup>2</sup>/E in the log rank test for survival data or tumour incidence data. *Biometrics*, **29**, 579-84.
- Peto R, Pike MC, Armitage P, et al (1977). Design and analysis of randomized clinical trials requiring prolonged observation of each patient II. Analysis and examples. *Br J Cancer*, **35**, 1-39.
- Radhakrishna S, Satagopan MC, Krishnaswami KV, et al (1980). A study of the accuracy and factors influencing accuracy of home addresses of patients obtained by registry clerks and address cards in four large towns in South India. *Tubercle*, **61**, 197-206.
- Rothman KJ (1978). Estimation of confidence limits for the cumulative probability of survival in life table analysis. *J Chronic Dis*, **31**, 557-60.
- Sankaranarayan R, Nair MK, Stanley G, et al (1995). Cervical cancer in Kerala: a hospital registry based study on survival and prognostic factors. *Br J Cancer*, **72**, 1039-42.
- Satagopan MC, Radhakrishna S, Krishnaswami KV, et al (1983). A novel system to obtain addresses of out-patients -assessment in routine clinic practice in Madras. *Ind J Tub*, **30**-39.
- Shanta V, Gajalakshmi CK (1989). Follow-up of cancer cases at the Cancer Institute, Madras. *Cancer Registry Abstract*, **3**, 6.
- Tallis GM, Leppard P, O'Neill TJ (1988). The analysis of survival data from central cancer registry with passive follow-up. *Stat Med*, **7**, 483-90.
- Tallis GM, Leppard P, O'Neill TJ (1993). Conditional life tables from registries with unidentified losses. *Stat Med*, **12**, 767-75.
- Varghese C, Ramdas K, et al (1991). Evaluation of reply-paid postcards as a cost effective follow-up strategy in Kerala. *Cancer Registry Abstract*, **5**, 12-4.