



SWAMINATHAN RAJARAMAN

An Evaluation of Survival of Cancer Patients
Based on Registry Data From Low or
Medium Resource Countries



ACADEMIC DISSERTATION

To be presented, with the permission of
the board of the School of Health Sciences of the University of Tampere,
for public discussion in the Small Auditorium of Building B,
School of Medicine of the University of Tampere,
Medisiinarinkatu 3, Tampere, on May 4th, 2012, at 12 o'clock.

UNIVERSITY OF TAMPERE

ACADEMIC DISSERTATION
University of Tampere, School of Health Sciences
Finland
Cancer Institute (WIA), Chennai
India
International Agency for Research on Cancer, Lyon
France

Supervised by
Professor emeritus Matti Hakama
University of Tampere
Finland

Reviewed by
Doctor of Medicine Cherian Varghese
University of Tampere
Finland
Doctor of Philosophy C. Ramesh
Kidwai Memorial Institute of Oncology
Bangalore, India

Copyright ©2012 Tampere University Press and the author

Distribution
Bookshop TAJU
P.O. Box 617
33014 University of Tampere
Finland

Tel. +358 40 190 9800
Fax +358 3 3551 7685
taju@uta.fi
www.uta.fi/taju
<http://granum.uta.fi>

Cover design by
Mikko Reinikka

Acta Universitatis Tamperensis 1724
ISBN 978-951-44-8785-9 (print)
ISSN-L 1455-1616
ISSN 1455-1616

Acta Electronica Universitatis Tamperensis 1193
ISBN 978-951-44-8786-6 (pdf)
ISSN 1456-954X
<http://acta.uta.fi>

CONTENTS

SUMMARY.....	6
LIST OF ORIGINAL PUBLICATIONS.....	9
ABBREVIATIONS.....	11
1. INTRODUCTION.....	12
1.1 Cancer registration principles and methods worldwide	12
1.1.1 Hospital-based cancer registry	12
1.1.2 Population-based cancer registry	14
1.2 Cancer survival studies from low or medium resource countries	18
2. REVIEW OF LITERATURE.....	21
2.1 Hospital-based cancer survival studies	21
2.2 Population-based cancer survival studies	22
2.3 Factors influencing population-based survival and comparisons	22
2.3.1 Host factors	22
2.3.2 Tumour related factors	24
2.3.3 Health care related factors	24
2.4 Data quality indices for population-based cancer survival study	25
2.5 Complete and incomplete follow up	26
2.6 Censoring: Potential withdrawals or loss to follow up	27
2.7 Bias due to type of loss to follow up	28
2.8 Ascertainment of non-randomness of censoring due to losses to follow up	29
3. AIMS OF THE STUDY.....	30
3.1 General	30
3.2 Specific	30

4. MATERIAL AND METHODS.....	31
4.1 Data sources	31
4.2 Classification of factors analysed	32
4.2.1 Residence	32
4.2.2 Clinical extent of disease	32
4.2.3 Cancer directed treatment	33
4.3 Methods	33
4.3.1 Follow up	33
4.3.2 Eliciting the determinants of non-random loss to follow up	35
4.3.3 Survival estimation	35
4.3.4 Relative survival and age-standardization	36
4.3.5 Loss adjusted rate	37
4.3.6 Elucidating bias in survival in the absence of active case follow-up	37
5. RESULTS.....	41
5.1 Descriptive statistics on characteristics of population based cancer registries from low or medium resource countries that had undertaken survival studies	41
5.2 Data quality indices for population-based survival study	45
5.3 Magnitude of loss to follow up in low or medium resource countries	50
5.4 Pattern of loss to follow up in Izmir, Turkey and Songkhla, Thailand	52
5.5 Determinants of non-random loss to follow up	54
5.5.1 Non-random loss to follow up – Population-based survival study	55
5.5.2 Non-random loss to follow up – Hospital-based survival study	56
5.6 Loss-adjusted survival studies in India	57
5.6.1 Loss-adjusted survival in population-based studies	58
5.6.2 Loss-adjusted survival in hospital-based studies	59

5.7 Risk of loss to follow up and loss adjusted survival in Thailand	60
5.8 Elucidating bias in survival under different assumptions on vital status	62
5.8.1 An application using Chennai PBCR data	63
6. DISCUSSION.....	68
6.1 Cancer survival differences in less developed and more developed countries	68
6.2 Data quality indices – Implications of lack of active follow up	71
6.3 Loss to follow up – Implications of loss-adjusted survival	76
6.4 Intra-country variation in cancer survival – Methodological implications	78
6.5 Survival differences – Implications of treatment or disease characteristics	79
6.6 A guide for choosing follow up methods	86
7. CONCLUSIONS.....	87
ACKNOWLEDGMENTS.....	89
REFERENCES.....	91
ORIGINAL PUBLICATIONS.....	99

SUMMARY

Cancer is a growing global health issue and many low or medium resource countries are ill-prepared to deal with the ever-increasing cancer burden owing to lack of well-developed surveillance systems. This needs an inter-disciplinary approach through international collaborations between low, middle and high income countries. Systematic reporting of cancer incidence and to some extent, cancer mortality, has been done periodically for many decades now. Unlike in well-developed countries, cancer survival however, is not routinely reported from low or medium resource countries. It required special and concerted efforts from multiple quarters to get reliable survival statistics.

Cancer survival generally refers to the lifetime of a person after the diagnosis. Population-based cancer survival data are essential for evaluating the development and distribution of and accessibility to cancer health services like treatment or screening. Since data from low or medium resource countries are beginning to surface in intermittent intervals, so have comparisons between well-developed and less-developed countries. This dissertation provides a stepwise methodological evaluation right from the conduct of survival study to the estimation of survival probability through empirical data from more than 25 registries in several low or medium resource countries with variable gross national income values. This is inevitable for a balanced interpretation of survival differences. The main material for study came from the SURVCAN database of the multinational study by the International Agency for Research on Cancer, Lyon, France, and is supplemented by several materials from India and Thailand.

The impact of variation in patient follow-up on survival statistics is undisputed. It could be due to inappropriate methods employed for getting vital status information: lack of active methods of follow up in the presence of sub-optimal mortality ascertainment or high

magnitude of loss to follow up by ineffective active follow up. In both instances, it is shown by empirical data that application of standard methodology results in systematic bias in the estimate of survival. If the losses are high and result in non-random censoring due to correlation with outcome, say death, it is a clear indicator to improve the follow up by vigorous active methods and to deviate from standard life table estimation of survival and resort to estimation of survival by differential loss-adjustment procedures explained through its determinants. The magnitude of bias varied between 1-4 percent units for population-based 5-year absolute survival and was larger between 2-7 percent units even for 3-year overall survival for hospital-based studies, for different cancers.

In a registry data environment that warranted the employment of active methods of follow up and the real losses to follow up did not exceed one in five cases, the bias induced in actuarial survival under different assumptions of vital status of cases due to inappropriate choice of follow up methods revealed the following: if only passive methods were employed, say for convenience or out of constraints, without any active follow up component, the bias induced in 5-year absolute survival estimates varied between 22-47 percent units for different cancers; when predominantly passive methods of follow up were employed with necessary active component, the bias ranged between 3-10 percent units; when follow up methods were totally by active methods but losses to follow up cases were excluded from analysis, the bias induced varied between 2-8 percent units for different cancers. This provides an objective index of bias resulting in over-estimation or under-estimation of survival in a low or medium resource country setting.

In these circumstances, age-standardized survival rates might adjust for the potential confounders and survival data by important prognostic factors like extent of disease may still appear plausible or consistent. But a systematic evaluation of bias in estimating survival due

to methodological problems and its suitable correction are mandatory before survival differences could be attributed to the varied development of treatment resources and/or disease characteristics in low or medium resources settings.

LIST OF ORIGINAL PUBLICATIONS

- I. Sankaranarayanan R, Swaminathan R, Brenner H, Chen K, Chia KS, Chen JG, Law SCK, Ahn YK, Xiang YB, Yeole BB, Shin HR, Shanta V, Woo ZH, Martin N, Sumitsawan Y, Sriplung H, Barboza AO, Eser S, Nene BM, Suwanrungruang K, Jayalekshmi P, Dikshit R, Wabinga H, Esteban DB, Laudico A, Bhurgri Y, Bah E and Hamdan NA (2010) : Cancer survival in Africa, Asia and Central America : a population-based study. *Lancet Oncol* 11(2):165-73.
- II. Swaminathan R, Rama R and Shanta V (2008a): Lack of active follow up of cancer patients in Chennai, India: implications for population-based survival estimates. *Bull World Health Organ* 86(7):509–15.
- III. Swaminathan R, Rama R and Shanta V (2008b): Childhood cancers in Chennai, India, 1990-2001: Incidence and survival. *Int J Cancer* 122(11):2607-11.
- IV. Sriamporn S, Swaminathan R, Parkin DM, Kamsa-Ard S and Hakama M (2004): Loss-adjusted survival of cervix cancer in Khon Kaen, Northeast Thailand. *Br J Cancer* 91(1):106-10.
- V. Ganesh B, Swaminathan R, Mathew A, Sankaranarayanan R and Hakama M (2011): Loss-adjusted hospital and population-based survival of cancer patients. In: *Cancer survival in Africa, Asia, the Caribbean and Central America*, pp 15-21. Eds. R Sankaranarayanan and R Swaminathan, IARC Scientific Publications No. 162, International Agency for Research on Cancer, Lyon.

VI. Sankaranarayanan R, Swaminathan R and Black RJ (1996): Global variations in cancer survival. Study Group on Cancer Survival in Developing Countries. *Cancer* 78(12):2461–4.

VII. Swaminathan R and Brenner H (2011): Statistical methods for cancer survival analysis. In: *Cancer survival in Africa, Asia, the Caribbean and Central America*, pp 7-13. Eds. R Sankaranarayanan and R Swaminathan, IARC Scientific Publications No. 162, International Agency for Research on Cancer, Lyon.

In the text, the above papers are referred by the roman numerals in brackets.

ABBREVIATIONS

ASRS	Age Standardized Relative Survival
CI	Confidence Interval
CI5C	Cancer Incidence in Five Continents
CONCORD	The short name for study on cancer survival in five continents
DCO	Death Certificate Only
DNA	Data Not Available
GNI	Gross National Income
HBCCR	Hospital Based Clinical Cancer Registry
HBCR	Hospital Based Cancer Registry
IACR	International Association of Cancer Registries
IARC	International Agency for Research on Cancer
LAR	Loss Adjusted Rate
LFU	Loss to follow-up
NFU	No Follow Up
NP	Not Published
OR	Odds Ratio
PBCR	Population Based Cancer Registry
SEER	Surveillance Epidemiology End Result
SURVCAN	The short name for the latest IARC multinational cancer survival study
USA	United States of America

1. INTRODUCTION

1.1 Cancer registration principles and methods worldwide

Reliable data on the magnitude of cancer problem are essential for monitoring the health of the community, assess the performance of the health care system and allow authorities to make informed decisions. Cancer registration may be defined as the process of continuing, systematic collection of data on the occurrence and characteristics of cancer with the purpose of helping to assess and control the impact of malignancies on the community. The cancer registry is the office or institution, which attempts to collect, store, analyze and interpret data on persons with cancer (Jensen et al., 1991). The potential source of reliable data has been the cancer registry, forming an essential part of any rational program on cancer control (Muir et al., 1985). Epidemiological research based on comprehensive cancer registration remains the most valid and efficient way to plan and evaluate cancer control activities. The value of a cancer registry is dependent on its quality and the extent to which it is used in research and health services planning. The usefulness of the data collected would be maximized by adopting uniform methods in all aspects of cancer registration. The data becomes useful for more and more purposes when they are accumulated over longer periods of time. The means of recording cancer cases by active or passive methods may be identical but a distinction is made between two major types of cancer registries: Hospital Based Cancer Registry (HBCR) and Population Based Cancer Registry (PBCR; Jensen et al., 1991).

1.1.1 Hospital-based cancer registry

A HBCR is concerned with the recording of information on all cancer patients seen in a single or group of hospitals, usually without the knowledge of the background population. In other words, all cancer patients attending the hospital(s), irrespective of the place or area they come from, are registered with an emphasis on clinical care and hospital administration. HBCRs present an opportunity to begin a documentation process to provide information on

clinical epidemiology of cancer (Valsecchi and Steliarova-Foucher, 2008). The establishment of HBCRs is historically rooted in the belief that individual patients are better served through the presence of a registry, since the registry will serve to ensure that patients return for follow-up examinations on a regular basis. The importance of a HBCR need not be stressed more than its existence is an indispensable requirement in the accreditation process of any cancer research programme of a hospital (Young, 1991). The HBCR ensures comparability of data between registries worldwide and over very long time period by adopting uniform classification of cancer diagnosis through standard international norms for disease coding (ICD-10, 1992, ICD-O, 2000). With the presence of a HBCR, case finding mechanisms are evolved so that the potential departments dealing with cancer cases and/or records are covered for accession and required information following a standard questionnaire format are collected either from patients (by direct interview with consent) and/or abstracted from records and/or by linkage through computers to serve as a repository of data on all cancer cases attending the institution (Young, 1991).

A HBCR is also central to monitoring the patient follow-up activity. This includes devising ways to record several patient contact particulars before the start of initial treatment as a prerequisite, to systematically update data on follow up visits of patients to the hospital and to initiate timely reminders for those patients who default through active methods like postal or telephone or other inquiries or approaches. Naturally, these activities make sure that a HBCR is an important source of data for any survival study (Jensen et al., 1991, Young, 1991). In most low or medium resource countries, a HBCR has usually been the starting point of cancer registration activity in a region before expansion into a population-based coverage (Valsecchi and Steliarova-Foucher, 2008). While all HBCRs under the National Cancer Registry Programme in India had been largely successful in achieving systematic and continuous registration of new cancers, the data on disease outcomes have largely been

deficient. This was mainly because of lack of strategy to develop follow up methods and documentation and integrate the same with HBCR activity. Apart from isolated reports on selected and small series of hospital patients, survival outcome based on large hospital series has been rare to come (Rao et al., 1998, Shanta et al., 2008). Chennai HBCR routinely publishes survival statistics on treated patients as part of the annual or biennial reports (Shanta et al., 2008). However, individual HBCRs in different major medical institutions have served as important data sources for PBCR thereby forming the nucleus of PBCR activity in the region.

1.1.2 Population-based cancer registry

The main objective of PBCR is to collect and classify information on all incident cancer cases occurring in a defined population, most specifically to a geographic area, in order to generate statistics on the occurrence of cancer in that population and to provide a framework for assessing and controlling the impact of cancer on the community (Jensen et al., 1991). The earliest population based cancer registry was commissioned in Hamburg, Germany in 1929, with emphasis on medical, scientific, public health and economic aspects through active form of registration of cases from multiple sources and subsequent comparison with death certificates as a follow up activity on a voluntary basis. The continuous recording of cancer cases by patient name began in Mecklenburg in 1937 signifying a methodological progress of eliminating multiple registrations and determining individual outcomes. Population-based cancer registry of New York State in USA was established in 1940 with compulsory notification of cancer cases. The Danish Cancer Registry, founded in 1942 is the oldest serving registry covering a national population (Jensen et al., 1991). Since then, this activity has gradually progressed and is currently well developed in high resource countries. In most of the well-developed countries, cancer has been declared as a notifiable disease and hence registration of incident cases is predominantly done by passive method. However,

population-based cancer registration is still in variable levels of development in low or medium resource countries

Unlike in the well-developed countries, cancer is not a notifiable disease in most low or medium resource countries and hence registration of incident cancer cases had been carried out predominantly by active methods as per the guidelines advocated by the International Agency for Research on Cancer (IARC) and the International Association of Cancer Registries (IACR). The location of registry is usually in the major cancer hospital with research facility in the region. The cancer registrars of the PBCR regularly visit multiple sources of data including major hospitals in government or public and private sectors, nursing homes, consultants, radiation centres, pathology laboratories, imaging centres, screening programmes, insurance firms and hospices, for data collection from patients by direct interview and/or from medical records or case listings or computer print-outs. A standardized form is used for collection of data on personal identification, disease, treatment and outcome variables. The mandatory data collected are as follows: patient identity (patient name and/or personal identity number, area of residence with particular emphasis on duration of stay of one or more years to avoid registering cases from a floating population, age at diagnosis and/or date of birth, sex) and disease related (incidence date, most valid basis of cancer diagnosis, cancer site and morphology, tumour behaviour and grade). Other data pertaining to the patient (socio-demographic, elements of socio-economic status, etc.), disease (clinical extent of disease and/or tumour stage) and treatment (received or not and/or type or modality, etc.) are collected as optional data, depending on the resources and availability. Data collection on deaths due to cancer, occurring in the region, is independently carried out as part of PBCR operations from vital statistics division as well as hospital death registers. It included data on deceased identity (name and/or personal identity number, age at death, sex, etc.) and death (all or cancer causes, date, place, etc.). The mortality data thus collected were

matched against all incident cancer cases in PBCR database through visual inspection of probable lists of similar pairs of listings manually or by electronic linkages. Data on all deaths, irrespective of the stated cause of death, were also utilized for this linkage to optimize the availability of mortality information on registered cancer cases. Matched cases were updated with death information in registry database. In a majority of PBCRs, unmatched deaths were traced back to hospitals for availability of more details on disease factors and registered accordingly. If no additional information is forthcoming, these deaths are registered in PBCR as cases on the basis of a death certificate only (DCO). Since cases are registered from multiple sources, elimination of duplicate notifications is done with utmost care. This is directly related to the quality of person identity data at registration. With the knowledge of background population that is giving rise to the cases, reports on incidence rates are published routinely. Even in low or medium resource countries, PBCRs have been extensively utilized in evaluating cancer screening and early detection programmes in the region (Swaminathan et al., 2009).

PBCR operations have been carried out in a systematic manner for many decades now even in low or medium resource countries (Table 1). The scientific publication series from the IARC, Lyon, France, titled Cancer incidence in five continents (CI5C) from volumes I to IX, constitute a compendium of cancer incidence statistics based on good quality data from cancer registries worldwide (Parkin et al., 2005, Curado et al., 2007).

Table 1: Current status of cancer registration and survival studies in low or medium resource countries by continent or region

Continent/ Region	Number of countries	Number of registries	Number of populations studied	Year of starting registration – Range	Conduct of population- based survival study		Conduct of survival study on hospital series	
					Countries	Year of 1 st publication	Countries	Year of 1 st publication
Africa	8	10	17	1953-1999	4	2003	1	1999
Asia	14	52	57	1960-2000	8	1995	6	1971
Caribbean	4	4	4	1958-1995	1	1996	0	-
Latin America	11	21	22	1958-2000	3	2006	3	1999

Table 1 shows the status of population based cancer registration in low or medium resource countries by continents or regions as included in volumes I to IX of CI5C series (Parkin et al., 2005, Curado et al., 2007). Cancer registration activity in Africa, Asia, the Caribbean and Latin America had commenced in late 1950s or early 1960s: Uganda, Kyadondo, in 1954; Israel in 1960; India, Mumbai (formerly Bombay) in 1962; Colombia, Cali, in 1967 (Parkin et al., 2005). New registries in low or medium resource countries have started their operations in mid or late 1990s and newer ones have been added to this list as recent as in early or mid-2000 (Curado et al., 2007). Collective or individual reports on cancer incidence and mortality have been published as a routine from many of the registries in low or medium resource countries continuously from time to time (Sierra et al., 1988, Laudico et al., 1989, National Cancer Registry Programme, 1992, Vatanasapt et al., 1993, Parkin et al., 2003, Shanta et al., 1994). Hospital and population based cancer registries, being the repositories of data on cancer cases collected in a systematic manner using standard methods, are generally regarded as important sources of information about cancer survival (Black et al., 1998c). Cancer registries could also serve as a novel alternative for long-term clinical trial follow up (Shi et al., 2010).

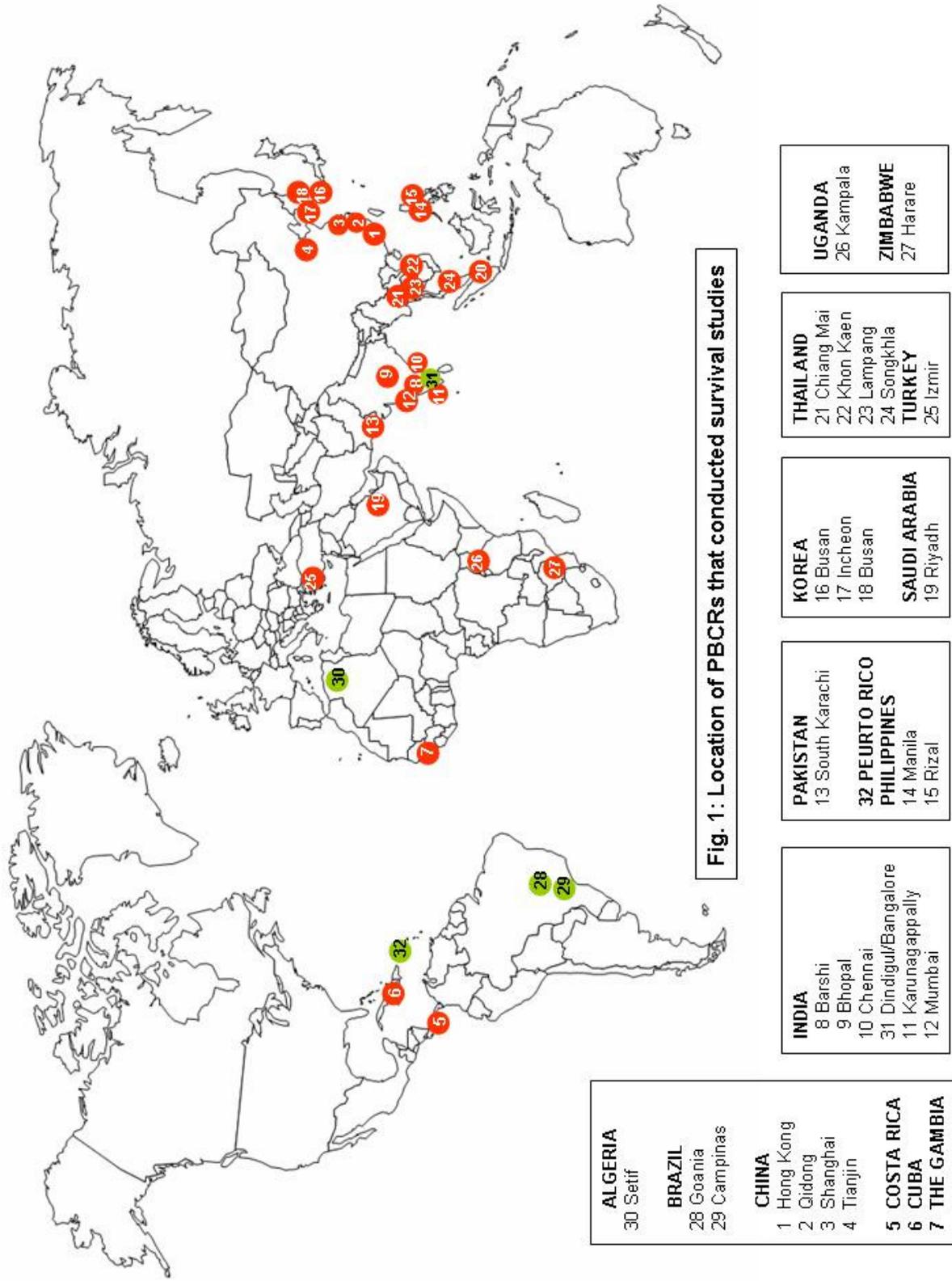
1.2 Cancer survival studies from low or medium resource countries

Long-term survival from cancer, such as surviving for five years or more after diagnosis, may reflect cure and is a positive sounding measure that can be used by planners, the public, doctors and patients to measure and discuss the outcome of cancer diagnosis and success of treatment. Survival analysis can be considered as a cohort study with a difference: here the length of follow up time is of greater interest than the occurrence of the event itself. Also, the rate of occurrence of the event is not constant over time and censored observations occur. Hence, special methods capable of dealing with such instances are necessary for undertaking survival analysis.

Hospital based cancer registries with high resolution database form the basis for survival studies on selected series of cancer cases. If survival study is part of randomized controlled clinical trials, it represents the gold standard for the evaluation of outcomes of treatment. Otherwise, it aims to provide information about the outcome of cancer directed treatment in particular settings in formulating hypotheses on the effectiveness of treatment modalities or outcomes and study of prognostic factors. On the other hand, survival rates calculated using PBCR data, with at least minimum information on all cancer cases in defined areas, would provide an objective index of the effectiveness of cancer care in the region concerned. However, cancer survival studies have generally been sparse from low (with per head gross national income (GNI) less than US\$ 2000) or medium (per head GNI between US\$ 2000 and US\$ 10,000) resource countries. Until early 1990s, there were only isolated reports of survival studies on hospital series based on locally available expertise and interests (Sankaranarayanan et al., 1998). Unlike in well developed countries, PBCRs in low or medium resource countries, with a long history of operations, have not been able to undertake survival studies routinely. Table 1 reveals that the earliest publications on cancer survival based on registry data in Africa, Asia, Caribbean and Latin America were brought

out between one and five decades after the commencement of registry operations. This points out to the general lack of good surveillance systems despite availability of long-standing cancer registration practices.

Figure 1 shows the map of location of 32 registries from 16 low or medium resource countries that had conducted population-based cancer survival studies till date. The first collaborative study on cancer survival in developing countries initiated by the IARC, Lyon, France, reported data from 10 registries in five low or medium resource countries (Sankaranarayanan et al., 1998). The second study by IARC called SURVCAN attracted participation from 26 registries in 13 low or medium resource countries. The following registries had participated in the first or second or both the studies: Hong Kong, Qidong, Shanghai and Tianjin from China; national registries of Costa Rica and The Gambia; Bangalore, Barshi, Bhopal, Chennai, Karunagappally and Mumbai registries from India; South Karachi from Pakistan; Manila and Rizal from Philippines; Busan, Incheon and Seoul from the Republic of Korea; Riyadh from Saudi Arabia; Chiang Mai, Khon Kaen, Lampang and Songkhla from Thailand; Izmir from Turkey; Kampala from Uganda and Harare from Zimbabwe. Singapore was an additional one from the high resource countries that also participated in the SURVCAN study (Sankaranarayanan and Swaminathan, 2011). The registries from low or medium resource countries that participated in the worldwide population study on cancer survival in five continents named CONCORD were Setif Wilaya registry from Algeria and Goania and Campinas registries from Brazil (Coleman et al., 2008). The national registry of Cuba had contributed data for all of the above studies.



2. REVIEW OF LITERATURE

2.1 Hospital-based cancer survival studies

Hospital-based cancer survival studies either form part of on-going clinical trials or aside studies on small numbers of selected cases with specific cancers pertaining to one institution or a small network of institutions. These studies usually involved selected cancer cases that completed at least one modality of cancer directed treatment following standard treatment protocols or by choice. The rationale of a randomized clinical trial is to eliminate even the effects of unknown confounders so that the only systematic differences are the treatments received (Black et al., 1998c). This approach is needed to establish the efficacy of the treatments. However, hospital-based studies on non-randomized settings are essential for the purposes of eliciting the effectiveness of treatment protocols that are being followed for different cancers in the hospital. In more developed countries, such patterns of care studies are mounted on HBCRs that have inherent systematic follow up procedures and transforming them into hospital-based clinical cancer registries (HBCCR) facilitating collection of high-resolution data on a continuous basis at least for selected major cancers like breast and cervix having good prognosis. Such an initiative is already in place under National Cancer Registry Programme in India.

The need to compute survival probability as an outcome measure of treatment was realized in the early 1970s in low or medium resource countries (Krishnamurthi et al., 1971). This was possible only because an effective follow up system was evolved and was made an integral function of the HBCR on a continuous basis to include major cancers. However, unlike in high resource countries, cancer survival studies based on hospital series were only sparingly available from low or medium resource countries for a long time from then: Pengsaa et al., 1989, Pavlovsky et al., 1992, Nair et al., 1993, Ganesh, 1995, Mathew, 1996,

Rao et al., 1998, to name a few. These studies dealt with varied aspects like estimating survival probability, eliciting prognostic factors and methodology.

2.2 Population-based cancer survival studies

In order to describe completely the experience of cancer in a population, it is necessary to know not only its incidence and mortality, but also the survival of cancer patients. Effectiveness of cancer services generally does not depend only on the efficacy of treatment but also on the context in which they are applied. Evaluating effectiveness requires estimation of survival in unselected groups of cancer patients and this is exactly what population-based survival study aims to provide (Black et al., 1998c). Unlike in high resource countries, the first report of survival studies from low or medium resource countries based on population based cancer registry series of all or selected incident cancers started emerging from the mid-1990s (Nandakumar et al., 1995, Sriamporn et al., 1995). These were the first results of the initiative taken by the IARC, Lyon, France, for conducting a multi-national collaborative study on cancer survival in developing countries in 1994 (Sankaranarayanan et al., 1998). Several such projects were undertaken later with the support of different international agencies by including new registries from low or medium resource countries in Africa, Asia, the Caribbean and Latin America (Coleman et al., 2008; Swaminathan et al., 2009, Sankaranarayanan and Swaminathan, 2011).

2.3 Factors influencing population-based survival and comparisons

2.3.1 Host factors

Age at diagnosis had emerged as an independent prognostic factor for many cancers from many registries with clear inverse relationships with survival (Sankaranarayanan et al., 1998). This could be either age may be associated with risk of dying due to particular cancer or dying due to other causes. This effect is often dealt by age-standardization of survival rates

by choosing appropriate standard methods (Brenner et al., 2004b) or populations (Black and Bashir, 1998a).

Sex of the patient is less commonly associated with variations in survival permitting combined data on men and women together for estimating survival. Such disparities in head and neck cancer survival are removed when adjusted for potentially confounding prognostic variables like life styles and treatment (Roberts et al., 2010). However, for some cancers like skin melanoma, sex of the patient is an independent risk factor for survival (Mervic et al., 2011) probably due to greater recognition of early symptoms.

Comorbid conditions experienced by cancer patients may vary substantially between registry populations. Comorbidity affects survival by presenting an additional source of risk of death, making it less likely that a patient will be offered curative treatment and if it is offered, less likely that the patient will be able to withstand the effects of treatment itself (Black et al., 1998c).

Socio-economic differences in survival have been reported for many cancers in Europe (Kogevinas, 1991, Cavalli-Björkman et al., 2011), the USA (Berg et al., 1977) and a few low or medium resources country populations (Nandakumar et al., 1995). Socio-economic disparities in diagnostic activity and management of large bowel cancers have been reported, which affect survival (Cavalli-Björkman et al., 2011). For almost all cancer sites, survival was consistently the highest for patients with the highest education and lowest for those with only basic education, showing that even in a potentially equitable society with high health care standards, like Finland, marked inequalities persist in cancer survival (Pokhrel et al., 2010, Cavalli-Björkman et al., 2011). When socio-economic conditions are grossly different between more-developed and less-developed countries, the inequalities in access to or development of cancer care are likely to be of particular significance in survival

studies from low or medium resource countries (Black et al., 1998c). However, many elements of socio-economic status are not routinely available for all cancer sites in registry data but are usually collected using extra efforts as a special study.

2.3.2 Tumour related factors

By convention, cancer registry data are aggregated within categories of anatomical sites defined by standard coding norms (ICD-10, 1992, ICD-O, 2000). One has to be wary of the differential distribution of subsites when making international comparisons on survival. This applies to variations in morphology types within the same cancer site. The stage or clinical extent of disease at diagnosis is the single most important factor determining survival. Therefore, variations in stage distributions of cancers in the populations being compared have a profound impact on survival. Variations in diagnostic technology could still prompt a measurement error in stage between more-developed and less-developed countries (Black et al., 1998c). But, when an inverse relationship between tumour stage or clinical extent of disease and survival were forthcoming, it would be reassuring of data quality on staging.

2.3.3 Health care related factors

There are numerous ways in which the development of or availability of or accessibility to screening or diagnostic or treatment facilities for cancer could influence cancer survival. Studies have shown that survival of cancer patients is prolonged after treatment in specialized cancer centres (Stiller, 1994). Karjalainen and Palva (1989) suggested that the use of a treatment protocol gave better results than that by the free choice of a physician in multiple myeloma. Markedly lesser survival from testicular cancer in Estonia compared to other regions in Europe is attributed to deficiencies in disease management like non-referral to oncologists after surgery, poor access to contemporary radiotherapy and general lack of coordination among specialised cancer centres (Aareleid et al., 2011). Survival differences

between Filipino-American patients and patients from Manila and Rizal registries for nine common cancers, higher in the former than latter, highlighted the importance of access to and utilisation of diagnostic and therapeutic facilities in low or medium resource countries (Redaniel et al., 2009). However, interpreting the differences as due to quality of care *per se* may be misleading in the absence of knowledge on selection of cases treated, which may explain the difference better (Black et al., 1998c).

2.4 Data quality indices for population-based cancer survival study

Variations in the quality of cancer registration data would complicate the interpretation of survival data based on routine cancer registry data (Hanai and Fujimoto, 1985). Only good quality result, fairly presented and with demonstrated use for cancer treatment and cancer control, would allow registries to continue and develop (Magrath and Litvak, 1993). Population-based cancer survival generally portrays a broader range of cancer control activities like screening or organization of treatment services (Black et al., 1998c). This is essentially because it is unbiased by selection of both, treated and untreated cases of specific or all incident cancers in the region across various sources of registration. Hence, completeness and accuracy of registration of incident cancer cases assume importance. If cases not registered represent a random sample of the total, there may not be any systematic bias introduced in survival results. However, the probability of getting registered is likely to be correlated with prognosis. Thus, frequency of cases excluded from survival analysis on any pretext, would have a marked impact on the survival estimate and hence have to be kept to the barest minimum. The measurable indices that would determine the population-based cancer survival data quality due to exclusion from analysis are summarized as follows (Sankaranarayanan et al., 1998, Sankaranarayanan and Swaminathan, 2011):

- Frequency of cases that were excluded from survival study owing to have been registered based on a death certificate only: dead cases with zero survival time and information on cancer known only from a death certificate.
- Frequency of cases that were excluded from survival study owing to lack of any follow up: cases with zero survival time and vital status unknown or lost to follow up (LFU) with zero survival time.

Other data quality biases concerning health related factors can also have an impact on the estimated survival. Over-diagnosis through population-based screening for prostate cancer almost certainly accounted for the changing incidence and corresponding survival in the USA (Howlader et al., 2011). With minimal exceptions, this phenomenon may not have any bearing on cancer survival statistics arising from most low or medium resource countries for any cancer site. Influences of diagnostic facilities on survival may be felt through improvements in sensitivity of accuracy, inducing stage migration and variations in stage-specific survival (Feinstein et al., 1985). However, there would not be any problem when survival comparisons were done for groups of patients with tumours of all stages together.

2.5 Complete and incomplete follow up

Adequate and complete follow up is an important prerequisite for any survival study. This had remained as the greatest impediment in the conduct of cancer survival studies in most low or medium resource countries. The reasons included less developed routine information systems (like registration, documentation, etc.), lack of unique linkages of incidence and mortality data and less efficient follow up methods. Complete follow-up is deemed to have been achieved when the vital status (alive or dead) at closing date of study or follow up is known for an individual. If not known, then the follow-up is incomplete. The frequency of cases with incomplete follow up is the most important data quality index for any survival

study. This can be explicitly measured in variable lengths of time from the index date, say date of first diagnosis, in an active follow up environment (Swaminathan et al., 1998). By active follow up, it is meant that the registry makes efforts voluntarily, to get follow up information on patients whose vital status is unknown, through personal (direct approach through person contact) or other (indirect approach without person contact) approaches (given in detail in section 4.3.1). However, in a passive follow up environment, information on deaths is routinely received either by-law or via an arrangement with the vital statistics division. Using this procedure, those patients for whom no information of death has been received are presumed to be “alive” until that point of time. The main requirement for this method to work efficiently is that there must be a high quality of registration of mortality data and unique data linkage possibilities, say personal identity number, which ensure the follow-up of cases to be complete with the exception of migration or rare losses. Active follow-up would supplement the latter in case of incomplete passive follow-up. A majority of registries in low or medium resource countries had resorted to active methods for follow up data collection on vital status owing to the absence of reliable health information system, especially cancer mortality registration. The magnitude of incomplete follow up instances occurring in survival studies from low or medium resource countries had been generally high up to 40% for different cancers (Swaminathan et al., 2002). The pattern of incomplete follow up information also displayed variation with most of that occurring within one year of diagnosis in most registries whereas it was after 5 years from diagnosis in very few (Suwanrungruang et al., 2011, Eser, 2011, Sriplung and Prechavittayakul, 2011, Garrote et al., 2011).

2.6 Censoring: Potential withdrawals or loss to follow up

Censoring is unique to lifetime data analysis. It occurs when exact lifetimes are known for only a portion of the individuals in the study and known to exceed certain values in the

remainder. It allows utilization of all information independent of the length of follow-up of an individual patient, so that, even recently diagnosed patients contribute to long-term survival (Black and Swaminathan, 1998b). It can occur in many ways. Censored cases are usually withdrawals, surviving at date of last follow-up: this date can be either individual for each patient or a common closing date for all patients. However, censorship in terms of losses to follow-up takes place if follow-up fails before this potential withdrawal. The date at which the individual is lost to follow up corresponds to the end of the period of observation. The available information on this date provides the status indicator (Chiang, 1968). There is a qualitative difference between these two groups of censored cases. It is therefore important to know the extent or magnitude, pattern and type of losses to follow up in any survival study.

2.7 Bias due to type of loss to follow up

When censoring occurs, either due to the termination of study at the closing date which is solely technical or due to every loss to follow-up that is unrelated to the outcome studied, say death, it is said to be random or non-informative censoring. When censoring occurs due to loss of follow-up which is related to death, it is known as non-random or informative censoring. Standard life table approaches for estimating survival probability such as the actuarial (Cutler and Ederer, 1958) or Kaplan-Meier (Kaplan and Meier, 1958) methods do not distinguish between these two groups of censorings and treat both of them alike. This may cause major bias since the estimates of absolute survival may be artificially raised if there are losses to follow up (Ganesh, 1995, Mathew, 1996). Little reliance can be placed on the estimated survival assuming random censoring when the magnitude of loss to follow-up is high (Swaminathan et al., 2002).

2.8 Ascertainment of non-randomness of censoring due to losses to follow up

Death due to any cause is the common end point for estimating overall survival in hospital based studies and absolute survival in population based studies. Hence, mortality ascertainment is vital for complete follow up. Death registration system is generally not well developed in most low or medium resource countries. This is reflected in the paucity of cancer mortality data from population based cancer registries published in CI5C series (Parkin et al, 2005). In this situation, it is reasonable to believe that losses to follow up may be due to or associated with this deficiency. It would be a good starting point to examine the factors that are associated with the risk of dying as possible determinants of losses to follow up (Ganesh, 1995). For this analysis, all cases censored before closure of the study and having had a follow-up of say, less than three or five years, constituted the loss to follow-up group (outcome) and the rest of the cases who are either dead or known to be alive on the closing date of follow-up are treated as censored (Ganesh, 1995).

3. AIMS OF THE STUDY

3.1 General

- To develop a realistic framework for the conduct, design and analysis of hospital-based and population-based cancer survival studies in low or medium resource countries.

3.2 Specific

- To estimate cancer survival rates in low or medium resource countries (I, VI)
- To evaluate whether the estimated survival and differences, if any, are subject to or reflecting,
 - Constraints in cancer registration and/or follow up methods (II),
 - Inappropriate choice of analytical methods for estimating survival (IV, V, VII)
and
 - Patient, disease, treatment or intervention characteristics (I, III)

4. MATERIAL AND METHODS

4.1 Data sources

The material for this study comprises data from both hospital and population-based cancer registries in several low or medium resource countries that formed the basis for the seven original publications cited in the appendix. The data, utilized in part or full in this dissertation, includes those from,

- (i) SURVCAN databases of the International Agency for Research on Cancer (IARC), Lyon, France, comprising 537,490 incident cases of 1-52 cancer sites or types in 27 PBCRs from 14 countries in Africa, Asia, the Caribbean and Central America, registered during 1990-2001 and followed through 2003, period varying for individual registries and other associated material of preceding years. Studies (I) and (VI) are fully based on SURVCAN databases and reported the summary results.
- (ii) Population-based cancer registry in Chennai, India, comprising 22,460 cases of 10 most common cancers and corresponding subtypes plus all tobacco related cancers registered during 1990-1999 and followed through 2001 (II)
- (iii) Population-based cancer registry in Chennai, India, comprising 1,274 cases of all childhood cancers, aged 0-14 years at diagnosis registered during 1990-2001 and followed through 2003 (III)
- (iv) Population-based cancer registry in Khon Kaen province, Thailand, comprising 601 cases of invasive cervical cancers registered during 1985-1990 and followed through 1995 (IV)
- (v) Hospital-based cancer registry comprising 336 new cases of invasive breast cancers diagnosed and treated at Tata Memorial Hospital, Mumbai, India, in 1985 and followed through 1988 (V).

4.2 Classification of factors analysed

4.2.1 Residence

Data on residential status assumes significance in both hospital and population-based registries. Unlike in a PBCR, there are no geographic limits restricting the registration of patients in a HBCR. However, the nature of the residential area is expected to have an impact on the follow up in the context of both registries. For the data from Mumbai HBCR, the residential status of patients was classified into two as those from Mumbai city and its neighbourhood compared to other farther districts. For the data from Khon Kaen PBCR in Thailand, the area (district) of residence was classified as Muang or surrounding districts and others. Both these classifications were based on the proximity or not to the super-specialty cancer hospitals in their respective regions.

4.2.2 Clinical extent of disease

Data on clinical extent of disease has been used as a viable surrogate for stage of disease for selected cancers in this dissertation. It has the greatest significance in correlating local factors with the estimated survival. This data is routinely available or collected by most registries in low or medium resource countries. The broad norms adopted in classifying this variable into four categories are as follows:

Localized: Tumour confined to the organ of origin, without invasion into the surrounding tissue or organ and without involvement of any regional or distant lymph nodes or organs;

Regional: Tumour not confined to the organ of origin with invasion into the surrounding tissue or organ, with or without the involvement of the regional lymph nodes and not involving or spread to the non-regional lymph nodes or organs;

Distant metastasis: Tumour involving or spread to the non-regional lymph nodes or distant organs;

Unknown: The above information is unknown.

For Mumbai hospital-based study, staging for breast cancer was possible with available documentation and was done following standard norms (Hermanek and Sobin, 1987)

4.2.3 Cancer directed treatment

Data on cancer directed treatment modality would be very helpful in explaining the differences in cancer survival in any setting. However, the availability of such data is limited in a PBCR than in HBCR. For the Mumbai HBCR study, data on treatment for female breast cancer was categorized as those receiving chemotherapy and not. For Khon Kaen PBCR study on cervix cancer from Thailand, data on treatment was categorized into two as those receiving any treatment and no treatment.

4.3 Methods

4.3.1 Follow up

The follow up data on vital status (alive or dead) of a patient is indispensable in the estimation of absolute or overall survival. With varying development of cancer information systems and capabilities of providing data on follow up, the registries worldwide have evolved several ways to achieve this purpose. The methods and approaches towards follow up of patients adopted by registries in low or medium resource countries contributing data in this study are summarized in Table 2.

Table 2: Methods of follow up data collection employed by registries classified by different approaches

Method of follow up	Direct approach through contact with/by patient or others	Indirect approach without any contact with patient or others
Active – Voluntary, always seeking for information	Registry personnel initiating the interview or contact with patient and/or others by any means: through postal or telephone or house visit or other inquiries for information on vital status	Registry personnel abstracting or linking information by repeated scrutiny of medical records/death certificates at sources of data (hospitals or vital statistics office) for updating or matching with registry incident cancer database
	<i>Action: A_E</i>	<i>Action: A_P</i>
Passive – Involuntary, mostly receiving information	Consultation or inquiries initiated by patient and/or others through postal or telephone or visit for consultation to hospital or through other means resulting in information on vital status	Automated linkage of registry incident cancer database with one or more databases on mortality, population, health, based on unique or available patient identity parameters (number, name, etc.)
	<i>Action: P_P</i>	<i>Action: P_E</i>

Table 2 summarizes the general characteristics for broadly classifying the various approaches undertaken by the registries in low or medium resource countries to obtain vital status (alive or dead) information, into active or passive methods of follow up. A registry is classified to have undertaken follow up entirely by active method if the vital status information with reference to a pre-specified date is almost completely obtained by actions A_E or A_P and very negligible from actions P_E or P_P . A registry is categorized to have employed predominantly active method of follow up if the vital status information with reference to a pre-specified date for a majority of cases is obtained by actions A_E or A_P and for the rest by actions P_E or P_P . A registry is categorized to have employed predominantly passive method of follow up if the vital status information with reference to a pre-specified date for a majority of cases is obtained by actions P_E or P_P , to some extent by action A_P and negligible by action A_E . A registry is classified to have undertaken follow up entirely by

passive method if the vital status information with reference to a pre-specified date for almost all cases is obtained by actions P_E or P_P , minimally from action A_P and not at all by action A_E . Thus, active methods are characterized by voluntary action by registry personnel towards personal contact with patient/others with the possibility of eliciting the time at which cases were lost to follow up with reference to a pre-specified date. Passive methods are mostly involuntary and devoid of personal contact with patient/others (Swaminathan et al., 2011).

4.3.2 Eliciting the determinants of non-random loss to follow up

Categorical factors (like age at diagnosis, sex, etc.), each with reference and subcategory levels, that have potential to influence either follow up (complete or loss to follow up) or survival (alive or dead) were first determined using logistic regression (risk expressed as odds ratios in univariate or multifactorial settings) or Cox proportional hazard model (Cox, 1972; risk expressed as hazard ratios in univariate or multifactorial settings using survival time information). The outcome event studied with respect to follow up was loss to follow up either at 3 years or 5 years from the index date which was the date of first diagnosis of cancer. A differential pattern of loss to follow up (LFU), either between factors or within subcategories of factors would indicate that such factors emerge as determinants of LFU with an association of non-random type (IV, V, VII).

4.3.3 Survival estimation

Death due to any cause was the end point studied for overall survival data series from both hospital and population-based cancer registries. Survival time was calculated as the duration between the date of first diagnosis of cancer and the date of death or date of loss to follow up or the closing date of follow up, whichever was earlier. Overall or absolute survival was calculated by actuarial method (Cutler and Ederer, 1958) unless otherwise specified. This method treated all censorings as random and potential withdrawals at closing date and losses

to follow up, in the same grouped annual interval of follow time were not distinguished. An example of the life table giving the calculations is given in Table 3.

Table 3: Illustration of the layout of the life table and calculation of cumulative survival probability by the actuarial method (VII)

Interval	Alive at beginning of interval	Last known alive during interval (censored)	No. of deaths during interval	Effective no. at risk	Conditional probability of death	Conditional probability of survival	Cumulative probability of survival (to end of interval)
$t_i - t_{i+1}$	n_i	w_i	d_i	$N_i = n_i - (w_i / 2)$	$q_i = d_i / N_i$	p_i	$P_{i+1} = \prod_{j=0}^i p_j$
0-1	3289	166	365	3206.0	0.114	0.886	0.886
1-2	2758	275	301	2620.5	0.115	0.885	0.784
2-3	2182	37	278	2163.5	0.128	0.872	0.683
3-4	1867	30	191	1852.0	0.103	0.897	0.613
4-5	1646	20	106	1636.0	0.065	0.935	0.573

4.3.4 Relative survival and age-standardization

Relative survival is defined as the ratio (Ederer et al., 1961) of observed to the expected survival in the general population of the same age and sex (Hakulinen, 1982) and was calculated to exclude the effect of competing causes of mortality and to facilitate survival comparisons between countries with different background mortalities. Expected survival probabilities for individual registries were estimated from country, age and sex specific life tables (Lopez et al., 2001). To account for the differences in the age structure of the cancer cases, relative survival was adjusted for age and reported as age-standardized relative survival (ASRS). For age standardization (Brenner and Gefeller, 2004a, Brenner et al., 2004b), the weights were defined as the ratio of the proportion of patients in the respective age group in the standard cancer population of estimated incident cancer cases from less developed countries together in the year 2002 (Ferlay et al., 2004) divided by the proportion

of patients in the respective age group in the study cancer population for every classified cancer site for every registry. Analyses were done using the publicly available macros (Brenner et al., 2002). All of the above methods and more are summarized in (VII).

4.3.5 Loss adjusted rate

Unlike traditional survival analysis which grouped withdrawals and losses together, loss adjusted rate (LAR) differentiated the two. Potential follow up time for all subjects was five years (three years) to estimate 5-year (3-year) loss adjusted survival. The choice of potential determinants or confounding factors and corresponding strata based on subcategories of chosen factors are made. Study subjects are classified into two main categories: those with complete follow up and those with loss to follow up. It is assumed that those lost to follow up in specific stratum have the same probability of death as others still remaining under observation and belonging to the same stratum. Accumulating over prognostic strata resulted in annual loss adjusted survival and cumulative loss-adjusted survival probabilities were calculated within the actuarial framework but different assumptions (IV, V). The method of calculating loss adjusted survival using logistic regression approach facilitated simultaneous adjustment of any number of determinants of loss to follow up and is a simplification of computational procedure to estimate expected deaths among those lost to follow up. The conditional probability of dying, conditional probability of surviving and the cumulative probability of surviving the current and subsequent annual intervals are done under the modified framework of generating life table (IV, V).

4.3.6 Elucidating bias in survival in the absence of active case follow-up

Different actuarial assumptions on the survival status of subjects were made during follow-up under active or passive or a mixture of both methods (II). Figure 2 gives the schematic representation of vital status of each subject under real circumstances and different actuarial

assumptions of follow up. Let y_0, y_1, \dots, y_4 represent the calendar years. The period y_0 to y_2 (say, 1990 to 1992) signifies the registration of cancer cases and y_0 to y_4 (say, 1990 to 1994) indicates the period of follow up. Subjects were designated as belonging to the following categories: *[A]* when they were matched with mortality data obtained by routine registry data linkage with official mortality statistics without any active follow-up; *[B]* when they could not be matched through routine registry data linkage with official mortality statistics and their death was ascertained through active follow-up; *[C]* when they were lost to follow-up but known to be alive until a specific date, with unknown survival status at the close of follow-up; and *[D]* when they had completed follow-up and were known to be alive on the closing date.

The follow up status was classified into four different case scenarios depending on the assumptions made, as follows:

Case 1: Purely passive follow-up only – Apart from cancer cases matched with deaths from vital statistics division, those not matched with official mortality data were presumed to be alive at the close of follow-up. In this scenario, subjects in category A were treated as having died on their respective dates of death, while subjects B, C, and D were assumed to be alive on the last day of follow-up in the analysis.

Case 2: Predominantly passive method with minimal active follow-up – Cases lost to follow-up were presumed to be alive on the last day of follow-up. In this scenario, subjects A and B were treated as having died on their respective dates of demise, while subjects C and D were treated as having been alive on the last day of follow-up.

Case 3: Purely active follow-up only – Cases lost to follow-up were censored on the last date on which their survival status was known. Under this case scenario, subjects A and B were treated as having died on their respective dates of demise; subjects in category D were treated as having been alive on the last day of follow-up, and subjects in category C were

treated as having been alive until a specific date and censored thereafter for the survival analysis, based on actuarial assumption.

Case 4: Predominantly active follow-up with minimal passive component – Cases lost to follow-up were excluded from the survival analysis. This resembles Case 3, excepting that subjects in category C were excluded from the survival analysis.

Absolute survival probability, also known as crude survival, was estimated through an actuarial approach. However, the assumptions made in this study differed from those normally made using the routine actuarial method (II).

5. RESULTS

5.1 Descriptive statistics on characteristics of population based cancer registries from low or medium resource countries that had undertaken survival studies

The study of characteristics of 28 population-based cancer registries in low or medium resource countries that have undertaken survival studies till date along with descriptive statistics reveal interesting differences on various parameters of cancer registration and survival (I, VI, Parkin et al., 2005, Curado et al., 2007). The registration of incident cancer cases has been done either entirely or predominantly by active methods in 19 (68%) registries and either entirely or predominantly by passive methods in the rest (32%). There were 15 (54%) registries covering only urban population, 10 (36%) covering both urban and rural populations and 3 (10%) were national registries.

The IARC evaluates the quality of data of cancer registries in a systematic and uniform manner by setting moderate to high standards for publication in the CI5C series. Out of 28 registries, data from 22 registries (78%) have been published in one or more volumes of CI5C without any reservation and 3 (11%) with some reservation. The common reservation expressed for publishing the incidence data of the latter three registries was the suspected excess degree of incompleteness in case finding or high proportion of cases registered on the basis of a death certificate only. The reasons for data from three remaining registries not included in CI5C series were because of either non-submission of data by registry (say Bhopal, India) for scrutiny by IARC or registry was of recent origin (say Dindigul from 2003) not covered by CI5C series so far. Thus, the data quality of cancer incidence data from a majority of registries has been good.

However, unlike morbidity data that was published in CI5C volumes for 25 registries, the cancer mortality data was not readily forthcoming in 9 (36%) registries while there were

reservations on completeness of cancer mortality data in 6 (24%) registries. Thus, it is clear that in a majority of the registries that had undertaken population-based cancer survival studies, the routine collection of cancer mortality data, though carried out as part of registry operations, remained constrained and the system was not fully developed or available.

In this background, it requires special efforts and separate resources on the part of most of the registries from low or medium resource countries to obtain complete data on follow up that remains the key to the conduct of any survival study on cancer. From Table 4, of the 28 population-based registries that have conducted survival studies based on a few hundreds to several thousands of cases of all or selected cases of 1-52 invasive cancers, 18 (64%) have used active follow up methods either entirely or predominantly to gather data on vital status of incident cancer cases. In the rest (36%), passive methods were employed for the purpose. The registries that practiced the same methods for cancer registration as well as follow up of incident cancer cases for vital status were 18 (64%): 15 were either entirely or predominantly by active methods and 3 were predominantly by passive methods. The registries that employed active methods for registration but switched to passive mode for follow up were 2 (7%; Tianjin, China and Incheon, Korea) while there were none vice-versa. Out of 15 registries for which mortality data was either not well developed or unavailable routinely or subject to severe incompleteness by CI5C series evaluation, 11 had augmented it and pursued follow up by employing active methods. It emerges very clearly that registries have put in extra efforts and allocated separate resources for undertaking the follow up of cancer cases for obtaining vital status information on incident cancer cases.

The extent of complete follow up at five years from the first diagnosis of cancer (index date) ranged between 76-100% for registries that employed passive follow up methods either entirely or predominantly. This figure ranged between 30-100% for registries that

pursued follow up entirely or predominantly by active methods. Thus, completeness of follow up was differential between population-based cancer registries using passive or active methods in low or medium resource countries. Inter-country and intra-country differences in frequency of cases with complete follow up existed even when the follow up methods were identical and approaches were the same (I, VI).

Table 4: Characteristics of cancer registries from low or medium resource countries that have undertaken population-based survival studies, 1990-2001*

COUNTRY/ Registry	Cancer registration		Follow up data [‡]		Survival study [‡]		IARC-CI5C volumes	
	Method	Population	Method	Complete %	Cancers	Cases	Incidence data	Mortality data
CHINA								
Hong Kong	P _P	U	P _E	100	45	110,190	Y	Y
Qidong	A _P	M	A _P	100	33	20,167	Y	Y _R
Shanghai	P _E	U	P _P	100	52	70,006	Y	Y
Tianjin	A _E	U	P _P	100	51	70,005	Y	N
COSTA RICA								
COSTA RICA	P _E	C	P _P	74-83	2	6,297	Y	Y
CUBA								
CUBA	P _E	C	P _P	95-99	17	8,150	Y _R	Y _R
THE GAMBIA								
THE GAMBIA	P _P	C	A _E	81-98	6	505	Y	N
INDIA								
Bangalore [#]	A _E	U	A _E	85-92	8	4,781	Y _R	N
Barshi	A _E	M	A _E	96-100	15	1,188	Y _R	N
Bhopal	A _E	U	A _E	100	16	1,863	N	N
Chennai	A _E	U	A _E	79-99	20	22,618	Y	Y
Dindigul [§]	A _E	M	A _E	86-96	12	1,045	N	N
Karunagappally	A _E	M	A _E	91-100	22	1,601	Y	Y
Mumbai	A _E	U	A _E	82-93	28	46,162	Y	Y
PAKISTAN								
South Karachi	A _E	U	A _E	67-76	4	677	Y	N
PHILIPPINES								
Manila	A _E	U	A _P	75-82	4	1,040	Y	N
Rizal	A _E	U	A _P	30	1	1,299	Y	N
SAUDI ARABIA								
Riyadh	P _P	U	P _P	80	1	298	N	N
SOUTH KOREA								
Busan	P _P	U	P _P	100	48	41,434	Y	Y
Incheon	A _E	U	P _P	100	42	20,563	Y	Y
Seoul	P _P	U	P _P	100	46	77,827	Y	Y
THAILAND								
Chiang Mai	A _E	M	A _E	59-100	36	7,276	Y	Y _R
Khon Kaen	A _P	M	A _P	40-83	13	2,253	Y	N
Lampang	P _E	M	P _P	96-100	40	11,195	Y	Y _R
Songkhla	A _E	M	A _E	50-86	36	6,589	Y	Y
TURKEY								
Izmir	A _E	M	A _E	79-98	12	4,381	Y	Y _R
UGANDA								
Kampala	A _E	M	A _E	47-87	15	1,916	Y	N
ZIMBABWE								
Harare	A _E	U	A _E	95-100	17	1,990	Y	Y _R

* Period varies for individual registries; [§] Year 2003; [#] 1982-1989; IARC-CI5C: International Agency for Research on Cancer - Cancer Incidence in Five Continents; A_E: Active method only; P_E: Passive method only; A_P: Predominantly active method; P_P: Predominantly passive method; U: Urban population only; M: Mixed covering rural + urban populations; C: Country on a whole; Y: Yes; Y_R: Yes with reservation; N: No; [‡] SURVCAN database

5.2 Data quality indices for population-based survival study

The comparison of descriptive statistics on data quality indices of population-based cancer survival for lung and breast cancers between selected registries classified on methods of cancer registration and follow up are given in Table 5 (SURVCAN).

Table 5: Data quality indices: Frequency of excluded cases expressed as proportion of death certificate only and no follow up cases for lung and breast cancers in selected registries from low or medium resource countries separately for passive and active methods of cancer registration and follow up, 1990-2001* (SURVCAN database)

Country/Registry	Lung				Breast			
	Total registered	DCO%	NFU%	Included %	Total registered	DCO%	NFU%	Included %
<u>Passive methods of registration and follow up</u>								
China, Shanghai	14,113	0.0	0.1	99.9	5,184	0.0	0.0	100.0
Costa Rica	DNA	DNA	DNA	DNA	2,854	2.5	11.2	86.3
Korea, Seoul	10,294	12.7	4.9	82.4	5,907	3.3	6.0	90.7
Thailand, Lampang	3,278	7.8	0.3	91.9	842	1.4	0.0	98.6
Range		0-13	0-5	82-100		0-3	0-11	86-100
<u>Active methods of registration and follow up</u>								
China, Qidong	3,303	0.3	0.4	99.3	669	0.3	2.4	97.3
India, Barshi	48	2.1	0.0	97.9	124	0.0	0.0	100.0
India, Mumbai	3,995	13.0	0.6	86.4	7751	5.2	0.7	94.1
Thailand, Songkhla	850	5.1	16.1	78.8	665	1.2	13.3	84.5
Range		0-13	0-16	79-99		0-5	0-13	85-100

DCO: Death certificate only; NFU: No follow up; DNA: Data not available

** Period varies for individual registries*

The figures for frequency of lung cancer cases included for survival analysis ranged between 82-100% for selected registries in Asia and Central America, employing passive methods and 79-99% for selected registries in Asia undertaking active methods (Table 5). The corresponding figures for breast cancer were 86-100% and 85-100% respectively. The range of DCOs for lung cancer was 0-13% in both groups of registries; the range for no follow up cases was 0-5% in registries with passive methods and 0-16% in registries with active methods for lung cancer. The range of DCOs for breast cancer was 0-3% in registries with passive methods and 0-5% in registries with active methods; the corresponding figures for no follow up cases were 0-11% and 0-13% respectively. The comparison of frequency of cases included for survival analysis out of total incident cases between the two groups of registries that pursued passive or active methods of case registration and follow up revealed minimal variation for lung and breast cancers (Table 5). This augurs well for the conduct of a population-based cancer survival study.

However, the frequency of cancer cases excluded from survival study (owing to being DCOs or no follow up (NFU) with zero survival time) and the frequency of cases with incomplete or loss to follow up (LFU) among cases included in survival study both have to be considered in unison to evaluate data quality in population-based survival study. There may be instances when one is minimal while the other is not. The following scenarios from real data present the different problems encountered pertaining to data quality indices and reiterate their possible impact on the population-based cancer survival.

Table 6: Frequency of cases registered as DCO or with lack of complete follow up for common cancers in Cuba, 1994-1995 followed through 1999*

Cancer/site	Total registered	Excluded from survival study		Included in analysis %	Complete follow up %	Incomplete follow up: % lost to follow up- years from diagnosis			
		DCO%	NFU%			<1	1-3	3-5	>5
Tongue	314	25.5	0.9	73.6	95.2	3.0	0.9	0.9	0.0
Mouth	355	25.9	0.9	73.2	93.4	3.1	1.9	1.2	0.4
Tonsil	82	32.9	1.2	65.9	98.1	0.0	0.0	1.9	0.0
Oropharynx	60	20.0	0.0	80.0	97.9	2.1	0.0	0.0	0.0
Colon	2491	49.7	0.2	50.1	99.2	0.2	0.2	0.3	0.1
Rectum	790	29.1	0.5	70.4	98.0	1.1	0.4	0.5	0.0
Anus	106	8.5	2.8	88.7	98.9	1.1	0.0	0.0	0.0
Larynx	1165	30.7	0.6	68.7	96.3	1.5	0.9	1.0	0.3
Breast	2929	25.6	0.3	74.1	97.0	0.9	0.3	1.4	0.4
Cervix uteri	1450	15.4	0.4	84.2	94.4	2.2	1.2	1.7	0.5
Urinary bladder	1182	29.5	1.2	69.3	97.4	1.2	0.0	1.0	0.4
Hodgkin lymphoma	320	40.3	1.6	58.1	97.4	1.6	0.5	0.5	0.0
Non Hodgkin lymphoma	771	39.7	0.1	60.2	97.7	0.6	0.6	1.1	0.0

DCO: Death certificate only; NFU: No follow up or lost to follow up with zero survival time;

** SURVCAN database*

Table 6 shows the frequencies of cases pertaining to all the data quality indices of population-based survival for selected cancers in Cuba during 1994-1995 and followed through 1999 (SURVCAN). Registration of cancer cases was carried out entirely by passive method. The frequency of cases registered as DCO ranges between 9-50% for different cancers. The very high figure may be the result of lack of active method of tracing back the cancer cases, first identified through a death certificate, to hospitals or to other sources of registration. On the other hand, the frequency of cases with zero survival time and vital status unknown were negligible ranging between 0-3%. In total, the frequency of cases included for survival analysis ranged between 50-89%. This is quite low and may or may not be a random sample or representative of the total incident cases. On the other hand, among the cases included for survival analysis, the complete follow up was achieved in 94-99%, which is adequate (Table 6). Follow up for vital status information was carried out predominantly by passive methods with minimal active component. Overall, despite good follow up, the resulting survival may not reflect the average outcome of respective cancers in the region owing to high degree of exclusion from the survival study thereby indicating high selection of cases.

The other kind of problem pertaining to inadequate follow up usually encountered in real data from low or medium resource countries is described in Table 7 (SURVCAN).

Table 7: Frequency of cases registered as DCO or with lack of complete follow up for common cancers in Khon Kaen, Thailand, 1993-1997 followed through 2000*

Cancer/site	Total registered	Excluded from survival study		Included in analysis %	Complete follow up %	Incomplete follow up: % lost to follow up- years from diagnosis			
		DCO%	NFU%			<1	1-3	3-5	>5
Lip	88	1.1	9.1	89.8	67.1	19.0	5.1	6.3	2.5
Tongue	57	0.0	5.3	94.7	77.8	20.3	1.9	0.0	0.0
Mouth	120	3.3	7.5	89.2	73.8	19.6	2.8	1.9	1.9
Nasopharynx	123	0.0	4.1	95.9	80.5	12.8	4.2	0.8	1.7
Colon	258	1.6	4.7	93.8	79.8	6.6	6.2	5.4	2.0
Rectum	143	0.0	2.8	97.2	80.6	10.1	6.5	0.7	2.1
Larynx	38	0.0	5.3	94.7	80.6	19.4	0.0	0.0	0.0
Breast	446	1.1	5.2	93.7	41.4	24.4	22.2	7.7	4.3
Ovary	230	0.9	9.6	89.6	47.1	28.6	12.6	8.7	3.0
Urinary bladder	114	0.0	9.6	90.4	41.7	35.9	10.7	8.7	3.0
Hodgkin lymphoma	31	0.0	6.5	93.5	37.9	27.6	17.2	6.9	10.4
Non Hodgkin lymphoma	191	0.0	12.0	88.0	53.0	28.0	10.1	6.0	2.9

DCO: Death certificate only; NFU: No follow up or lost to follow up with zero survival time;

** SURVCAN database*

Table 7 describes both the two data quality indices pertaining to cancer registration and follow up as observed in survival data from Khon Kaen, Thailand, in 1993-1997 and followed through 2000. Both cancer registration and follow up were carried out predominantly by active methods. The frequency of cases registered as DCO was minimal ranging between 0-3% for different cancers. This possibly is due to trace-back procedures by active methods adopted by the registry. However, the frequency of cases excluded from survival study with zero survival time and vital status unknown was higher ranging between 3-12%. In total, the frequency of cases included for survival analysis ranged between 88-97%, which is satisfactory. Among the cases included for survival analysis, the complete follow up was achieved only in 38-81% for different cancers. Overall, despite minimal exclusions from the study, the resulting absolute survival is not likely to reflect the average outcome of respective cancers in the region owing to the very high degree of incomplete or losses to follow up (19-62%), especially due to high losses to follow up within the first year from diagnosis (7-36%). These data quality indices, with the exception of DCOs, are applicable to hospital-based survival studies also.

5.3 Magnitude of loss to follow up in low or medium resource countries

The magnitude of loss to follow up observed in population-based cancer registries from low or medium resource countries for major cancers are shown in Table 8 (SURVCAN).

Table 8: Frequency (%) of cases lost to follow up for major cancers in registries from low or medium resource countries in Africa, Asia, the Caribbean and Central America, 1990-2001* (SURVCAN database)

Country/Registry	Mouth	Stomach	Rectum	Larynx	Lung	Breast	Cervix	Ovary	Bladder
China, Shanghai	2.5	0.7	1.1	0.5	0.4	1.7	1.3	1.0	1.2
Costa Rica	DNA	DNA	DNA	DNA	DNA	24.2	32.6	DNA	DNA
Cuba	6.6	DNA	2.0	3.7	DNA	3.0	5.6	DNA	2.6
The Gambia	DNA	12.8	DNA	DNA	19.4	8.2	4.5	DNA	DNA
India, Barshi	3.6	4.4	DNA	3.0	4.3	4.0	0.7	DNA	DNA
India, Chennai	26.5	10.6	DNA	17.2	10.4	23.3	38.0	22.6	17.0
India, Karunagappally	3.3	0.0	0.0	2.2	1.4	1.1	0.6	0.0	2.6
India, Mumbai	20.4	11.6	20.9	20.2	10.1	25.1	25.8	15.6	23.5
Pakistan, S. Karachi	28.8	DNA	DNA	DNA	DNA	DNA	DNA	DNA	DNA
Philippines, Manila	DNA	DNA	DNA	DNA	DNA	22.4	23.6	DNA	DNA
Philippines, Rizal	DNA	DNA	DNA	DNA	DNA	73.7	DNA	DNA	DNA
Saudi Arabia, Riyadh	DNA	DNA	DNA	DNA	DNA	21.8	DNA	DNA	DNA
Thailand, Chiang Mai	29.7	17.4	25.3	26.8	11.6	34.9	29.8	30.9	27.8
Thailand, Khon Kaen	26.2	DNA	19.4	19.4	DNA	58.6	62.1	52.9	58.3
Thailand, Lampang	0.8	0.3	0.8	1.2	0.2	1.1	1.4	0.5	1.9
Thailand, Songkhla	31.6	24.5	34.5	33.3	19.6	48.8	45.9	51.4	42.6
Turkey, Izmir	DNA	DNA	32.5	35.6	DNA	39.3	48.4	38.7	38.2
Uganda, Kampala	DNA	25.6	20.0	DNA	12.8	40.7	47.3	27.3	DNA
Zimbabwe, Harare	DNA	1.6	3.0	5.5	0.0	3.1	1.5	0.0	2.3

* Period varies for individual registries; DNA: Data not available

The magnitude of loss to follow up (LFU) at the end or closing date of follow up is observed to be very high in a majority of registries that have undertaken population-based survival study from low or medium resource countries (Table 8). The range of LFU across major cancer sites exceeded 20% for most cancer sites in Costa Rica (24-33%); South Karachi, Pakistan (29%); Manila, Philippines (22-24%); Rizal, Philippines (74%); Riyadh, Saudi Arabia (22%); Chiang Mai, Thailand (12-35%); Khon Kaen, Thailand (19-62%); Songkhla, Thailand (20-51%); Izmir, Turkey (33-48%) and Kampala, Uganda (13-47%). It was intermediate in The Gambia (5-19%); Chennai, India (11-38%) and Mumbai, India (10-26%). Excepting Costa Rica and Riyadh registries, all others had adopted active methods of follow up. It is likely that the high magnitude of loss to follow up would have a significant impact on the estimation of absolute survival in these registries. The range of LFU across major cancer sites was very minimal in Shanghai, China (0-3%); Cuba (2-7%) and Lampang, Thailand (0-2%) pursuing passive methods of follow up and Barshi, India (1-4%); Karunagappally, India (0-3%) and Harare, Zimbabwe (0-6%) adopting active methods of follow up (Table 8).

5.4 Pattern of loss to follow up in Izmir, Turkey and Songkhla, Thailand

The frequency of cancer cases that are lost to follow up being higher for major cancers in most registries from low or medium resource countries, it would be necessary to study the pattern of losses to follow up and its possible impact on the survival rate. The pattern of losses to follow up (LFU) is determined by the frequency of losses at variable lengths of time from the diagnosis of cancer. This is generally measurable only in registries adopting active methods of follow up.

Table 9: Frequency (%) and pattern of loss to follow up for selected cancers in Izmir, Turkey (1995-1997) and Songkhla, Thailand (1990-1999)*

Cancer site	LFU% at closing date	Loss to follow up (LFU%) – Years from diagnosis				LFU% at 5-years
		<1 year	1-3 years	3-5 years	>5 years	
Turkey, Izmir (1995-1997)						
Rectum	32.5	16.0	2.9	2.0	11.6	21.9
Larynx	35.6	9.8	1.7	1.5	22.6	13.0
Breast	39.3	10.9	2.3	1.8	24.3	15.0
Cervix	48.4	10.6	3.7	2.4	31.7	16.7
Ovary	38.7	10.9	1.3	2.6	23.9	14.8
Bladder	38.2	13.3	2.7	1.2	21.0	17.2
Range	32.5-48.4	9.8-16.0	1.3-2.9	1.2-2.6	11.6-31.7	13.0-21.9
Thailand, Songkhla (1990-1999)						
Rectum	34.5	10.3	9.9	7.9	6.4	18.1
Larynx	33.3	13.0	8.0	4.3	8.0	25.4
Breast	48.8	12.1	9.8	11.0	15.8	32.9
Cervix	45.9	13.2	12.1	8.6	12.1	33.8
Ovary	51.4	16.2	19.7	8.1	7.5	43.9
Bladder	42.6	11.0	11.6	9.7	10.3	32.3
Range	33.3-51.4	10.3-16.2	8.0-19.7	4.3-11.0	6.4-15.8	18.1-43.9

*LFU: Loss to follow up; * SURVCAN database*

Table 9 shows the two different patterns of losses to follow up encountered in Songkhla registry in Thailand and Izmir registry in Turkey (SURVCAN). In Songkhla registry, the highest number of losses to follow up ranging between 10-16% had occurred at less than one year from diagnosis for most cancers; it ranged between 8-20% at 1-3 years from diagnosis, 4-11% at 3 to 5 years from diagnosis and 6-16% after 5 years from diagnosis. The LFU% ranged between 33-51% at the end of follow up and between 18-44% at 5 years

from diagnosis for different cancer sites (Table 9). Here, the impact of this pattern of LFU on the estimation of 5-year absolute survival would be significant.

On the other hand, in Izmir registry, the highest number of losses to follow up ranging between 12-32% had occurred at 5 or more years from diagnosis for most cancers; it ranged between 10-16% at less than 1 year from diagnosis and ranged between 1-4% at 1 to 5 years from diagnosis (Table 9). The LFU% that ranged between 33-48% at the end of follow up got reduced between 13-22% at 5 years from diagnosis for different cancer sites. Here, the impact of this pattern of LFU on the estimation of 5-year absolute survival would be minimal. Hence, the pattern of losses to follow up is also an important contributing factor along with the magnitude of losses to follow up in the estimation of absolute survival. However, these are usually not estimable when registries undertake follow up based on purely passive methods.

5.5 Determinants of non-random loss to follow up

It is evident that a high proportion of censored cases in survival studies from low or medium resource countries are due to losses to follow up rather than being potential withdrawals. Also, the patterns of LFU were variable. Hence, it is essential to elicit the type of loss to follow up and its determinants. This is applicable to both hospital-based and population-based survival studies.

5.5.1 Non-random loss to follow up – Population-based survival study

Determinants of loss to follow up had been elicited by computing the risk of loss to follow up expressed as hazard ratio in multifactorial setting.

Table 10: Determinants of loss to follow up and non-random censoring by Cox proportional hazards model, female breast cancer (n=273), Mumbai PBCR, 1992-1994 followed through 1999, (VII)

Determinants of loss to followup	Loss to follow up		Relative hazard of loss to follow up [§]	
	No.	%	Hazard ratio	95% CI
Age at diagnosis				
<=44 years	53	6.9	1.00	-
45-54	75	10.4	1.63	1.14-2.31*
55-64	89	16.2	2.47	1.76-3.47*
65-74	47	13.7	2.25	1.52-3.33*
75+	9	7.4	1.26	0.62-2.56
Extent of disease				
Localised	129	14.6	1.00	-
Regional	98	8.2	0.54	0.41-0.71*
Distant metastasis	1	0.4	0.05	0.01-0.35*
Unknown	45	26.9	2.40	1.70-3.38*

*% computed to respective categories; CI: Confidence interval; * p<=0.05; § Each factor is adjusted for the other in the table*

Table 10 shows the proportion and risk (hazard ratio) of loss to follow up at 5 years from diagnosis with respect to the prognostic factors like age at diagnosis and clinical extent of disease, experienced in Mumbai population-based cancer registry for invasive breast cancer cases registered during 1992-1994 and followed through 1999 (VII). The proportion of loss to follow up between sub-categories of age at diagnosis and extent of disease varied widely and ranged between 7-16% and 0-27% respectively. A differential risk pattern of loss to follow up was observed between the sub-categories of prognostic factors. One-two folds higher risk of loss to follow up was forthcoming for all age groups more than 45 years

compared to less than 45 years, which was statistically significant. The risk of loss to follow up was 46-95% less among those with regional or distant metastatic disease compared to localized category. The risk was reversed and more than two-fold higher for cases with unknown extent of disease. All of these were statistically significant. Both age at diagnosis and extent of disease have emerged as determinants of loss to follow up in Mumbai PBCR. The differential risk pattern of loss to follow up with respect to each of the prognostic factors for survival indicate that the censoring due to losses to follow up are likely to be non-random.

5.5.2 Non-random loss to follow up – Hospital-based survival study

Table 11: Number and proportion (%) of patients and losses at 3-years and risk (odds ratio) of loss to follow up with 95% confidence interval by patient characteristics among female breast cancer patients diagnosed in Tata Memorial Hospital, Mumbai India, in 1985 and followed through 1988 (V)

Patient characteristics	Patients (n=336)		Lost to follow up (n=80; 24%)		Odds ratio (95% CI)
	Number	% ^a	Number	% ^b	
<u>Age at diagnosis</u>					
≤ 44 years	101	30	22	22	1.0*
45-54	117	35	29	25	1.2 (0.6-2.3)
55-64	77	23	19	25	1.2 (0.6-2.5)
65+ years	41	12	10	24	1.0 (0.5-2.9)
<u>Residence</u> (Mumbai city)					
Mumbai+neighborhood	169	50	26	15	1.0*
Other districts	167	50	54	32	2.6 (1.5-4.6) [§]
<u>Stage of disease</u> (TNM summary)					
I	29	9	4	14	1.0*
II	160	48	30	19	1.5 (0.5-5.6)
III	126	37	40	32	2.9 (0.9-10.6)
IV	21	6	6	29	2.5 (0.5-12.9)
<u>Treatment</u>					
With chemotherapy	194	58	42	22	1.0*
Without chemotherapy	142	42	38	27	1.3 (0.8-2.3)

* Reference category; [§]: p=0.05; CI: confidence interval;

^a percentage to total breast cancer cases; ^b percentage to total cases in respective categories

Table 11 examines the distribution and risk of losses to follow up (expressed as odds ratio) in Mumbai hospital-based survival study of newly diagnosed and treated female breast cancer cases with respect to more number of prognostic factors for survival like age at diagnosis, tumour stage, residence (proximity to base hospital) and cancer directed treatment (V). The differences in the risk estimates of loss to follow up between various categories of these factors compared to the corresponding reference sub-categories reflect the degree of association or non-randomness in the data on follow up. If differential pattern of risk of loss to follow up was observed with respect to any factor and was statistically significant, then that factor is a determinant of loss to follow up resulting in non-random censoring. The frequency of loss to follow up at 3 years from diagnosis was 24% but showed wide variation among sub-categories of every prognostic factor: LFU ranged between 22-25% for age at diagnosis, 15-32% for residence area, 14-32% for tumour stage and 22-27% for treatment groups. The risk of loss to follow up was elevated in excess of two-fold for residents living outside Mumbai city compared to those living in the neighbourhood and was statistically significant. A differential risk of loss to follow up was observed with respect to other prognostic factors also though not statistically significant. The risk was 2-3 folds higher for advanced stages of disease compared to early stage; 30% excess risk was forthcoming among those not treated with chemotherapy compared to those who received it and 20% elevated risk among those aged 45-64 years compared to less than 45 years. It is clear that differential loss to follow up exists and the censoring induced by these losses to follow up is likely to be non-random in this hospital-based study (V).

5.6 Loss-adjusted survival studies in India

The risk differences in table 10 and table 11 correlate with the risk of death as well (V, VII). This signifies a differential losses to follow up indicating the presence of bias and

this would impact the estimation of survival probability by the traditional approach using actuarial assumption. This observation on follow up data on female breast cancer from both hospital and population-based studies from Mumbai, India, stresses the need for estimating overall or absolute survival by specific loss adjusted methods that accounts for the differential losses to follow up.

5.6.1 Loss-adjusted survival in population-based studies

Table 12: Number of incident cases, proportion (%) lost to follow up and comparison of 5-year absolute survival with and without loss-adjustment for top ranking cancers in population based cancer registry, Chennai, during 1990-96 and followed through 2001 (V)

Cancer site/type	Number of incident cases	Lost to follow up %	5-year survival%		Absolute difference in survival
			Actuarial	With loss adjustment	
Cervix	3134	21.8	52.1	50.4	1.7
Breast	1923	20.7	39.5	39.1	0.4
Stomach	1845	8.0	9.4	8.7	0.7
Oesophagus	1403	6.7	7.7	7.5	0.2
Lung	1237	7.8	8.2	8.1	0.1
Mouth	1202	11.6	30.1	29.1	1.0
Lymphomas	768	11.5	26.5	25.6	0.9
Tongue	670	13.0	20.2	18.9	1.3
Leukaemias	668	8.2	19.8	19.2	0.6
Ovary	521	24.0	25.7	24.2	1.5

The magnitude of losses to follow up and the ensuing absolute survival at 5-years from diagnosis with actuarial assumption and specific adjustment for non-random losses to follow up have been extended to cover major cancers in Chennai population-based cancer registry, India, registered during 1990-1996 and followed through 2001 in Table 12 (V). The losses to follow up ranged between 7-24% for different cancers and were ascertained to be

non-random with respect to one or more of prognostic factors like age at diagnosis, sex, education, extent of disease and treatment (V). The 5-year absolute survival estimated using the standard actuarial assumption ranged between 8% for lung cancer and 52% for cervical cancer. The corresponding figures for loss adjusted survival were 8% and 50% respectively. The absolute differences in survival between these two approaches of estimation were minimal and did not exceed 2% for any cancer. Also, the loss-adjusted survival was seen to be lesser than the corresponding actuarial survival for all cancers (Table 12).

5.6.2 Loss-adjusted survival in hospital-based studies

Table 13: Comparison of 3-year survival without loss adjustment by actuarial assumption, stepwise loss-adjustment of factors using stratified method and loss adjustment using all factors together by logistic regression for all female breast cancer patients diagnosed in Tata Memorial Hospital, Mumbai India, in 1985 and followed through 1988 (V)

Loss adjustment of factors	3-year survival%
Without loss adjustment and using actuarial assumption only	61.2
Loss adjustment done by stratification	
Residential status	59.5
Residential status and age at diagnosis	59.2
Residential status, age at diagnosis and stage of disease	57.4
Residential status, age at diagnosis, stage of disease and treatment	56.5
Loss adjustment done by logistic regression	
Residential status, age at diagnosis, stage of disease and treatment	54.5

The effect of stepwise loss adjustment by its determinants like residence, age at diagnosis, tumour stage and treatment, in order, on overall cumulative survival compared to the standard actuarial method in a hospital-based survival study on female breast cancer from

Mumbai, India, is given in Table 13 (V). The 3-year overall survival of all breast cancers together by actuarial method was 61.2%. The corresponding 3-year survival, loss-adjusted for all the four determinants together were 54.5%. The absolute difference in survival was 6.7%. Stepwise loss adjustment by stratification from single to all four factors revealed a decreasing 3-year survival after addition of every factor. This showed that there was a marked effect on the survival estimate by loss-adjustment in the hospital based study.

5.7 Risk of loss to follow up and loss adjusted survival in Thailand

The effect of differential loss-adjustment on 5-year absolute survival based on more number of prognostic factors as determinants of loss to follow up, taken together as well as adjusted for one another, in a population-based survival study of 601 invasive cervical cancer cases in Khon Kaen province in Thailand, during 1985-1990, is given in Table 14 (IV). The overall loss to follow up at five years from date of first diagnosis of cancer was 27.6%. There was wide variation between sub-categories of different factors under consideration: 24-30% for age at diagnosis; 22-38% for tumour stage; 22-40% for treatment status and 27-28% for residence area. Each of the factors emerged as a potential prognostic factor for survival with a statistically significant association with death: 2-4 fold increased risk of dying, expressed as odds ratio, with increasing age at diagnosis; 2-5 fold excess risk of dying for other factors. The risk of loss to follow up was 2-fold higher for cases not treated compared to those received treatment and was statistically significant. This shows that there is differential non-random loss to follow up with respect to the prognostic factors to emerge as determinants.

Table 14: Number of cases, proportion and odds ratio (with 95% CI) of death and loss to follow up (LFU) at 5-years from index date, 5-year cumulative absolute and loss adjusted survival of factors studied for cervical cancer in Khon Kaen province, Thailand, 1985-1990 followed through 1995 (IV)

Factors	Number of cases	Proportion at 5 years from index date		Odds ratio (OR) with 95% CI		5-year absolute survival %	
		LFU%	Dead%	LFU-OR ^a	Dead-OR ^a	Actuarial	Loss adjusted ^a
All cervix cancers	601	27.6	36.4			56.8	54.7
Age at diagnosis							
<40	122	27.9	24.6	1.0	1.0	71.0	68.1
40-49	194	24.2	35.1	0.9 (0.5-1.5)	1.5 (0.8-2.7)	59.7	57.8
50-59	158	29.1	36.7	1.2 (0.7-2.0)	2.0 (1.1-3.7)	55.7	53.4
60+	127	29.9	49.6	1.2 (0.7-2.1)	3.5 (1.8-6.9)	38.8	37.6
Stage of disease							
Stage I	93	23.7	20.4	1.0	1.0	77.1	74.6
Stage II	134	28.4	29.1	1.2 (0.7-2.2)	1.8 (0.9-3.6)	65.5	63.0
Stage III/IV	222	21.6	53.6	0.8 (0.4-1.5)	5.0 (2.7-9.5)	39.0	38.2
Unknown	152	37.5	27.6	1.4 (0.7-2.6)	1.5 (0.7-3.1)	63.3	59.8
Treatment							
Received	428	22.4	36.2	1.0	1.0	59.2	57.5
Not received	173	39.9	37.0	2.0 (1.3-3.1)	2.0 (1.2-3.4)	49.9	47.5
Residence district							
Muang + neighbourhood	274	28.1	32.9	1.0	1.0	61.1	58.6
Other districts	327	26.9	39.4	0.8 (0.6-1.2)	1.5 (1.0-2.3)	53.2	51.4

^a Each factor adjusted for other factors in the table; CI: confidence interval

The 5-year absolute survival of all cases together by actuarial method was 56.8%. The corresponding 5-year survival, after differential loss-adjustment by taking all the four determinants together, was 54.7%. The absolute difference between these two was minimal (2.1%). An inverse relationship between survival and age at diagnosis or known stages of disease was observed: a decreasing survival with increasing age at diagnosis or stage of

disease. The maximum absolute difference in survival by actuarial and loss-adjustment did not exceed 3.5 units for any factor (Table 14).

5.8 Elucidating bias in survival estimate under different assumptions on vital status

The problems of high magnitude of loss to follow up, the pattern of high losses within the first year of follow up and the non-random type of losses to follow up necessitated the computation of loss-adjusted survival in both hospital and population-based survival studies in low or medium resource countries. It was clear that such loss-adjusted survival was lesser than actuarial survival for almost all cancer sites. This suggests that the patients who were lost to follow up had higher mortality than assumed in actuarial assumption of eliciting survival rate. In most low or medium resource countries, optimal mortality ascertainment is directly dependent on the methods adopted to obtain the data. Hence, it is important to study the effects of active or passive methods of ascertainment of vital status on the estimated actuarial survival. In other words, when the health information systems are generally not well developed, it is vital to elucidate the bias, if any, resulting from absolute survival estimates in the absence of active follow up and when different assumptions are made regarding the vital status (alive/dead) of cancer patients.

5.8.1 An application using Chennai PBCR data

Table 15: Number of cases variably classified as alive or dead under different assumptions of follow up by cancer site, Chennai PBCR, 1990-1999 (II)

Cancer/Site	Number of cases by vital status included for analysis									
	Total	Passive follow up only			Passive + active follow up					
		Case 1		Case 2	Case 3		Case 4			
		Dead	Presumed alive at closing date	Dead	Presumed alive at closing date	Dead	Alive	LFU	Dead	Alive
Lip	86	17	69	46	40	46	10	30	46	10
Tongue	988	371	617	693	295	693	54	241	693	54
Oral cavity	1662	528	1134	1052	610	1052	169	441	1052	169
Tonsil	250	107	143	214	36	214	16	20	214	16
Hypopharynx	1017	421	596	833	184	833	59	125	833	59
Oesophagus	2016	1028	988	1759	257	1759	59	198	1759	59
Stomach	2681	1392	1289	2277	404	2277	120	284	2277	120
Pancreas	328	190	138	291	37	291	23	14	291	23
Larynx	722	290	432	456	266	456	142	124	456	142
Lung	1806	1069	737	1574	232	1574	45	187	1574	45
Breast	3067	875	2192	1489	1578	1489	862	716	1489	862
Cervix	4438	1131	3307	1874	2564	1874	878	1686	1874	878
Ovary	808	321	487	487	321	487	138	183	487	138
Urinary bladder	442	172	270	305	137	305	62	75	305	62
Hodgkin lymphoma	298	92	206	171	127	171	74	53	171	74
Non Hodgkin lymphoma	868	383	485	602	266	602	130	136	602	130
Lymphoid Leukaemia	433	197	236	323	110	323	49	61	323	49
Myeloid Leukaemia	465	277	188	365	100	365	35	65	365	35
Leukaemia unspecified	85	56	29	69	16	69	5	11	69	5

PBCR: Population Based Cancer Registry; LFU: Lost to Follow Up

Case 1: Passive follow up only without any active follow up with cancer cases not matched with official mortality database presumed to be alive on the closing date of follow up.

Case 2: Passive + Active follow up with lost to follow up cases presumed alive on the closing date.

Case 3: Passive + Active follow up with lost to follow up cases censored at the last known date.

Case 4: Passive + Active follow up with lost to follow up cases excluded from survival analysis.

Table 15 gives the distribution of cases of major cancers by vital status, variably classified as alive or dead, based on different assumptions made on each case following the method adopted for getting the outcome data on follow up (II). In the example of cervix cancer, there were a total of 4,438 cases included for survival analysis. Of these, in reality, there were 1,131 (25.5%) deaths matched from the official mortality data from vital statistics division, 743 (16.7%) cases were known to be dead by active follow up methods (by actions A_E or A_P as in Table 2 in section 4.3.1) yielding a total of 1874 (42.2%) deaths; 878 (19.8%) cases were known to be alive on the closing date of follow up on December 31, 2001; for the rest of 1686 (38.0%) cases, alive or dead status was not known on this date as they had been lost to follow up at variable times between 1990 and 2001. The vital status of cases gets transformed when different assumptions are made.

If we assume that only passive follow up was adopted and no active follow up was pursued in Chennai, then we know death information on 1131 cases only, instead of 1874 deaths in reality. By rule, under passive follow up environment, if death information is not forthcoming, such cases are presumed to be alive. So, 743 deaths obtained by active methods of follow up plus 878 cases actually alive on closing date plus 1686 cases whose vital status was actually unknown at closing date will all be erroneously assumed to be alive on closing date (Case 1-purely passive method of follow up).

Suppose we assume that there is a moderately developed health information system functioning locally which can correctly identify all deaths occurring in the region and minimal active follow up with impersonal approach is adopted for follow up (by actions A_P as in Table 2 in section 4.3.1), then at the most, 1874 deaths will be known. By rule, under passive follow up environment, 878 cases actually alive on closing date plus 1686 cases that

were actually lost to follow up at variable intervals between 1990 and 2001 would all be assumed to be alive on the closing date (Case 2-predominantly passive method of follow up).

Suppose we assume the situation as it exists, then there would be 1874 deaths of cervix cancer cases, 878 cases alive on closing date and 1686 cases lost to follow up before closing date (Case 3-active method of follow up). Suppose we decide to exclude the cases, on whom, the alive or dead status was not known on closing date, then the analysis will include only 2752 cases: 1874 deaths and 878 alive cases, instead of 4438 (Case 4-predominantly active method of follow up).

There is variable impact caused by above scenarios on different cancers based on the availability of mortality data and extent or magnitude of losses to follow up. It is evident that the absolute survival estimated under different assumptions as stated above will also be different for different cancers. It becomes important to elucidate the bias arising out of such different misclassification of vital status (alive or dead) of cases in the analysis for major cancers.

Table 16: 5-year absolute actuarial survival% estimated under different assumptions on the vital status by method of follow up, Chennai PBCR 1990-1999 cases followed through 2001 (II)

Cancer site or type	Methods of follow up and different vital status assumptions			
	Passive follow up		Passive + Active follow up	
	All cases not matched with official mortality database are presumed as alive at the closing date	All cases lost to follow up are presumed as alive at closing date	All cases lost to follow up are censored at last known vital status date	Excluding all cases lost to follow up within five years of diagnosis from analysis
	Case 1	Case 2	Case 3	Case 4
Lip	79.5	44.6	40.7	39.5
Tongue	62.1	29.2	19.4	15.4
Oral cavity	68.5	37.1	30.5	26.4
Tonsil	58.5	17.2	13.7	10.8
Hypopharynx	59.2	20.0	12.5	9.6
Esophagus	48.9	12.9	6.9	5.0
Stomach	47.9	15.0	8.6	5.6
Pancreas	41.8	10.9	7.9	6.5
Larynx	59.0	35.1	30.7	28.4
Lung	40.8	13.2	6.5	4.2
Breast	71.6	51.5	43.7	39.6
Cervix	75.5	59.0	54.0	49.4
Ovary	60.1	39.5	27.4	21.1
Urinary Bladder	61.3	31.0	23.2	20.0
Hodgkin's lymphoma	69.1	42.6	39.4	35.9
Non Hodgkin's lymphoma	55.6	29.7	21.6	16.8
Lymphoid Leukemia	54.3	26.5	23.8	15.5
Myeloid leukemia	40.4	21.5	14.7	10.9
Leukaemia unspecified	32.9	17.8	10.9	6.2

The five year absolute survival estimated by actuarial methods under different assumptions on the vital status of the cases based on corresponding assumption on the follow up environment or methods of follow up as totally passive or active or a mixture of both using Chennai population-based cancer registry data during 1990-1999 are given in Table 16 (II). The survival estimated by following case 1 was the highest and by following case 4 was the lowest for all cancers. Case 3, which treats vital status of subjects as is, is taken as standard. The absolute difference in estimated survival by case 1, case 2 and case 4, compared to case 3 represents the bias. In the absence of active follow up (case 1), 5-year absolute survival was estimated to be higher by 22% (in leukaemia unspecified) to 47% (in hypopharyngeal cancer) than when cases were actively followed and were lost to follow up at a known point in time (case 3). The bias ranged between 3 (for pancreas) and 10 (for tongue) percent units for case 2 vs. case 3. When follow up methods were totally by active methods but losses to follow up cases were excluded from analysis, the bias induced varied between 2-8 percent units for different cancers. The more losses to follow up the greater are the uncertainty and potential for bias, in the actuarial estimate. Cases 2 and 4 represent the two extremes of a survival spectrum, with the actuarial estimate assuming random withdrawal (case 3) falling in between. The absolute differences in 5-year survival between cases 2 and 4 were substantial for cancers of the tongue (13.8%) and ovary (18.4%).

6. DISCUSSION

6.1 Cancer survival differences in less developed and more developed countries

Documentation of cancer cases has to be perceived as a substantial part of cancer control programme than as a bureaucracy component (Valsecchi and Foucher, 2008). The fundamental step in carrying out an end-result study is to ensure adequate complete follow up. If the vital status (alive or dead) of all the cases included in the study is known at the closing date, excepting for rare losses and random drop outs as experienced in well developed country setting, the estimation of survival probability by standard life table methods is straightforward and unbiased. Every registry contemplating a survival study strives hard to achieve optimal level of complete follow up. However, in low or medium resource countries, it is difficult to obtain complete follow up information for all patients for various reasons: less-developed health information systems, especially mortality, that limit the data linkage possibilities; restrictions in active data collection through personal contact owing to data confidentiality agreement with multiple data sources; requirement of additional resources encompassing expertise, personnel, funding, etc. Hence, a high magnitude of loss to follow up and a marked variation in the completeness of follow up data between registries is on the expected lines from most registries in less developed countries. Information from all cases is used, including cases whose follow up ends due to closure of the study and those lost to follow up before closure in the estimation of survival. This differentiation has an impact on the survival statistics from these registries and should be borne in mind while interpreting survival differences between any two registries between or within countries or regions, especially in low or medium resource setting.

Table 17: 5-year Age Standardised Relative Survival (ASRS%; 0-74 years) of major cancers by country in low or medium resource settings (I):^c Comparison with US-SEER White (1996-2002),^a EUROCARE-4 (1995-99),^b and Singapore (1993-1997)^c survival

Country (Registries)	5-year Age standardized relative survival (0-74 years of age)									
	Median (Minimum-Maximum) of values if more than one registry are contributing									
	Lung	Stomach	Large bowel	Breast	Cervix	Bladder	Ovary	Larynx	Oral cavity	Tongue
US-SEER White*	15	22	65	90	71	82	44	66	60	58
EUROCARE-4 ^s	12	24	54	82	67	71	37	63	49	45
Singapore [†]	09	27	52	76	66	72	62	66	49	44
China (Hong Kong (U)/ Qidong (R)/ Shanghai (U)/ Tianjin (U))	21 (6-32)	39 (20-44)	44 (36-63)	82 (58-90)	67 (48-79)	78 (47-81)	56 (35-67)	68 (49-73)	67 (44-71)	67 (64-68)
Costa Rica [†]	NA	NA	NA	70	53	NA	NA	NA	NA	NA
Cuba ^{† c}	NA	NA	21	70	56	69	NA	58	44	39
The Gambia [†]	20	03	04	12	22	NA	NA	NA	NA	NA
India (Barshi (R)/ Bhopal (U)/ Chennai (U)/ Karunagappally (M)/ Mumbai (U))	7 (1-13)	6 (4-15)	28 (6-31)	52 (31-54)	46 (34-60)	39 (10-48)	25 (19-29)	28 (16-38)	37 (26-45)	23 (12-30)
Pakistan (Karachi (U))	NA	NA	NA	NA	NA	NA	NA	NA	41	39
Philippines (Rizal(U)/Manila(U))	NA	NA	40	47 (40-55)	37	NA	NA	NA	NA	NA
Saudi Arabia (Riyadh (U))	NA	NA	NA	64	NA	NA	NA	NA	NA	NA
South Korea (Busan (U)/ Incheon (U)/ Seoul (U))	20 (14-21)	49 (46-50)	60 (57-64)	79 (78-81)	79 (76-79)	76 (72-79)	59 (54-62)	62 (59-76)	52 (48-54)	53 (52-60)
Thailand (Chiang Mai (M)/ Khon Kaen (M)/ Lampang (M)/ Songkhla (M))	9 (5-12)	12 (9-20)	35 (31-44)	63 (57-66)	61 (54-63)	48 (33-60)	47 (45-58)	36 (26-49)	36 (22-42)	32 (24-33)
Turkey (Izmir (M))	NA	NA	52	77	63	71	60	71	NA	NA
Uganda (Kyadondo (M))	00	00	08	46	13	NA	09	NA	NA	NA
Zimbabwe ^c (Harare (U))	11	23	28	58	39	37	34	30	NA	NA

* 5-year relative survival; ^s area weighted 5-year relative survival among adults; [#] includes anus; NA: Not available; [†] National registries; R: Rural; U: Urban; M: Mixed; ^a Ries et al., 2006; ^b Sant et al., 2009; ^c SURVCAN database

The 5-year age standardized (0-74 years) relative survival for major cancers by country from medium or low resource settings are compared with corresponding survival in regions from high resource settings in Table 17. The countries from low or medium resource settings can be grouped into three tiers based on observed survival. Survival was the highest in Hong Kong (in China), South Korea and Turkey, where health services are well developed with advanced diagnostic and treatment centres and high per head Gross National Income (GNI) values. Survival was intermediate in Costa Rica, mainland China, Thailand, India, Pakistan, Philippines and Zimbabwe, where cancer health services are moderately developed with diagnostic and treatment facilities centred in and around urban cities and with medium per head GNI. Survival was the lowest in The Gambia and Uganda, with poorly developed health services, as indicated by limited availability of cancer diagnostic and treatment facilities and with very low per head GNI. The 5-year survival reported for most low or medium resource countries in SURVCAN study were lower than that reported for the white patients in the United States Surveillance Epidemiology End Results (US-SEER) program for most cancers (Ries et al., 2006). The country-weighted 5-year relative survival data from 23 European countries (Sant et al., 2009) and 5-year ASRS in Singapore (Chia, 2011) were on par with one or more registries from China, South Korea and Turkey for cancers of the breast, cervix, large bowel, head and neck. The level of development of health services and their efficiency in providing early diagnosis, treatment and clinical follow up care can have a profound effect on cancer survival. However, a meaningful interpretation is possible only after taking into account the differences in data quality between registries, especially when they are from a wide range of economic development levels.

6.2 Data quality indices – Implications of lack of active follow up

The fact that 25 out of 28 registries from low or medium resource countries that have undertaken population-based survival studies had contributed data to IARC CI5C publication series (Parkin et al., 2005, Curado et al., 2007) at one time or other, stands testimony to the data quality on cancer incidence. However, in 15 registries, the mortality data were not published as they were not routinely available or were included with considerable reservation. Thus, the important data quality issue in a survival study is achieving adequate follow up to get vital status data whether the patient is alive or dead at the end of the study. In a low or medium resource setting, with demonstrated less-developed mortality registration systems, achieving adequate complete follow up is possible only if registries undertook special efforts by evolving a variety of active methods suiting the conditions. Survival reported by most registries that pursued follow up entirely by active methods, tended to reflect the average outcome from the different cancers studied, keeping with the advanced stages at presentation, standards of health care development in their regions, inequities in accessibility to services especially cancer directed treatment and compliance to it and minimal or no cancer screening facilities. Interestingly, the countries that achieved the highest survival in this study have pursued follow up of cases predominantly by passive means with minimal active components.

Box plots have been employed to examine the relationship, if any, between the estimated 5-year ASRS and four categories of registries classified based on methods (A_E , A_P , P_E and P_P according to Table 4) adopted for follow up data collection for vital status. The published five-year age-standardized relative survival (ASRS) percent values for cancers of the breast (Figure 3a) and cervix (Figure 3b) were utilized from registries that contributed data registered during 1990-2001 and period varying for individual registries (Sankaranarayanan and Swaminathan, 2011). The median, quartiles and range of ASRS (0-74

years) values showed a gradual ascendancy from entirely active to entirely passive methods of follow up. This phenomenon was true for most cancers with high lethality as well (Sankaranarayanan and Swaminathan, 2011). This suggests a possible methodological problem of follow up, especially in the ascertainment of deaths, as demonstrated in Chennai registry data, resulting in substantial bias in the actuarial survival estimate under standard assumptions.

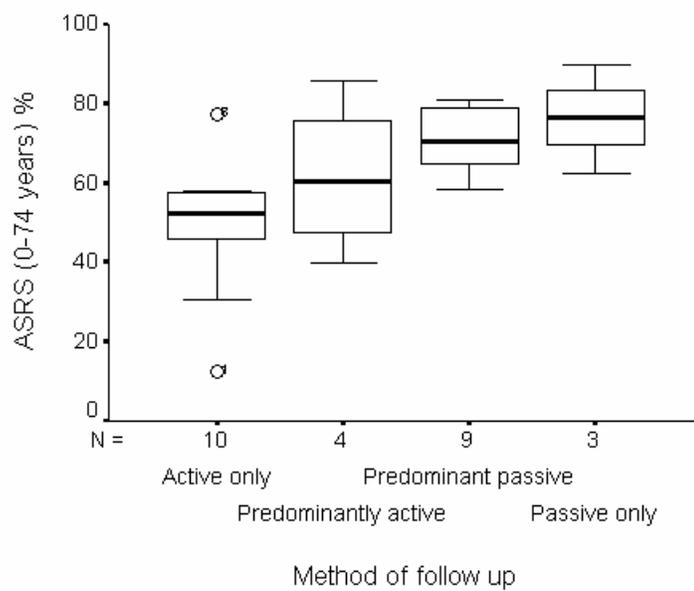


Figure 3a: Breast cancer 5-year Age Standardised Relative Survival (ASRS; 0-74 years) by classified methods of follow up in 26 registries, 1990-2001

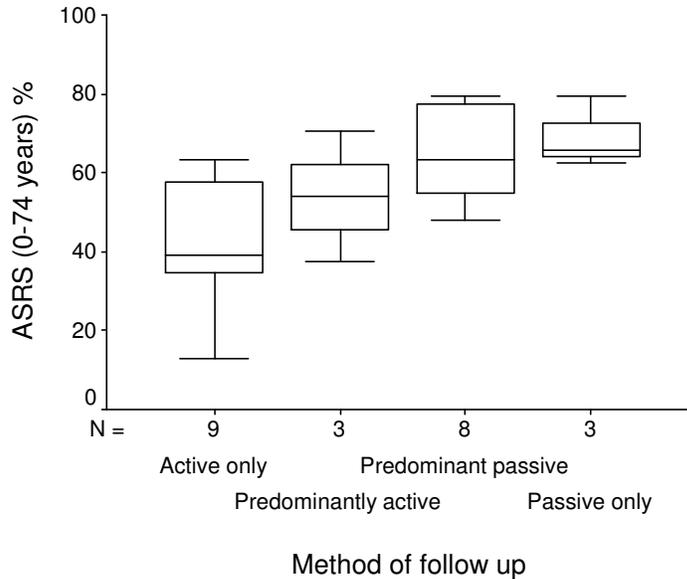


Figure 3b: Cervix cancer 5-year Age Standardised Relative Survival (ASRS; 0-74 years) by classified methods of follow up in 23 registries, 1990-2001

High level of completeness of both cancer incidence and ascertainment of mortality data are important prerequisites for valid cancer survival estimates and when such completeness cannot be assured, survival rates and their comparisons should be carefully interpreted. Even modest levels of under-registration of deaths may lead to severe overestimation of long-term cancer survival estimates (Brenner and Hakulinen, 2009). Mortality ascertainment will be sub-optimal in a passive follow up environment if the data linkages between mortality and incident cancer registry databases are not based on a unique personal or national identification number and not backed by a sound death registration system. Such deficiencies result in incomplete follow up. This is even more accentuated by the fact that death registration is generally done based on place of occurrence of death and not necessarily on usual place of residence. The registries generally have access only to official mortality data of the region covered by the registries. This discounts the possibility of knowledge of deaths of cancer patients occurring outside of the registry coverage area if deaths obtained from linkages alone are accounted. This purportedly suggests re-examining

the definition of location of death in health services research and death registration system. Hence, such defective linkages effectively means that the extent of incompleteness in follow-up is unknown, especially mortality and such dead cases would have been erroneously classified as alive at the closing date of follow up. Therefore, the reported survival from some of the registries in the study may clearly be over estimated.

Aside studies have been carried out in the past to estimate the completeness of follow up, especially mortality, in a passive environment by extra active follow up for selected cancers, which revealed missing of deaths through routine linkages (Berrino et al., 1995). The present study has extended the elucidation of the bias in the estimation of absolute survival under different assumptions on vital status of patients depending on purely or predominantly active or passive methods of follow up for major cancers using Chennai registry data from India. This analogy is applied to cover the age-standardized relative survival rates for the same cancers in Table 18.

Table 18: Absolute increase in Age Standardised Relative Survival (0-74 years)% between treating loss to follow up as is and assuming them as alive at closing date in the absence of active follow up in Chennai, India, 1990-1999

Cancer site/type	% lost to follow up: Years from diagnosis				Dead%	Age Standardised Relative Survival (0-74 years) %		
	Up to 5 years	<1 yr	1-3 yrs	3-5 yrs		With loss to follow up (LFU)	LFU as alive on closure	Absolute increase
Lip	9.3	7.0	2.3	0.0	53.5	49.0	53.6	4.6
Tongue	16.9	13.1	2.6	1.2	70.1	23.4	34.1	10.7
Oral cavity	14.4	10.3	2.2	1.8	63.3	36.7	44.0	7.4
Tonsil	5.6	4.8	0.8	0.0	85.6	15.6	20.2	4.6
Hypopharynx	10.0	9.0	0.7	0.3	81.9	15.0	23.4	8.4
Oesophagus	8.0	6.7	0.9	0.4	87.3	8.6	15.4	6.8
Stomach	9.1	7.3	0.9	0.9	84.9	10.3	17.7	7.4
Pancreas	3.7	3.1	0.3	0.3	88.7	8.7	11.9	3.2
Larynx	7.8	6.7	0.8	0.3	63.2	38.0	43.4	5.4
Lung	8.9	8.0	0.7	0.3	87.2	7.1	14.6	7.5
Breast	17.4	12.4	2.9	2.0	48.5	47.7	56.2	8.5
Cervix	17.2	11.0	3.7	2.5	42.2	59.6	64.9	5.3
Ovary	20.8	14.7	4.6	1.4	60.3	28.5	41.8	13.3
Bladder	12.7	10.9	1.6	0.2	69.0	32.0	41.9	9.9
Hodgkin's	9.1	6.4	1.7	1.0	57.4	37.8	41.4	3.6
Non Hodgkin's	12.8	10.9	1.3	0.6	69.4	23.2	32.2	9.0
Lymphoid leukaemia	9.5	2.8	3.2	3.5	74.6	16.4	20.6	4.2
Myeloid leukaemia	10.3	8.6	1.3	0.4	78.5	14.9	21.9	7.0
Leukaemia unspecified	10.6	10.5	0.0	0.0	81.2	11.2	17.1	5.9

The implications of lack of active follow up on population based survival in Chennai, India, under two assumptions on the survival status of patients for major cancers registered during 1990-1999 and followed through 2001 in Chennai, India, are given in Table 18. It shows the 5-year ASRS% values by assuming that all loss to follow up cases in reality, as alive on 31st December 2001 (Case 2 in methods section 4.3.6; figure 2) and by treating all losses to follow up cases with actuarial assumption (Case 3 in methods section 4.3.6; figure 2). An upward bias ranging between 3-13% under 5-21% of losses to follow up for different cancers was detected. Extending the same analogy to other registries that pursued predominantly passive methods of follow up to get vital status information, if the losses to follow up did not exceed one in five and they were not correlated with survival, the over-estimation of 5-year ASRS would have conformed to the upper limit of 13%. The more the losses, the higher would be the bias. The bias would be much more if the mortality ascertainment by passive follow up was poorer than anticipated. Therefore, it is imperative for registries from low or medium resource countries to evolve suitable methods of active follow up before embarking on survival studies.

6.3 Loss to follow up – Implications of loss-adjusted survival

The magnitude of the bias depends on the extent of loss to follow up: the lesser the loss implies lesser bias. This holds true, if the losses are common and correlated with the patient prognosis or survival. In most low- or medium-resource countries, such losses are common due to deficiencies in health infrastructure and recording of health statistics, especially mortality. The losses are also likely to be related to the patient's chance of survival. Differential losses may be encountered for a variety of prognostic factors: low socio-economic status may be responsible for lack of continuous patient surveillance; stage or extent of disease and/or age at diagnosis may be related to the motivation of follow-up, are a few examples. Furthermore, the direction of bias is unpredictable and gets accentuated in

case of non-random nature of losses to follow up or informative censoring. In such instances, it is desirable to investigate deviation from randomness of censoring using the information on all available factors associated with survival.

The losses to follow up have been of high magnitude and common to both hospital and population-based registries that pursued a variety of active methods of follow up. The patterns of losses have also been different: high losses within a year or 1-5 years or more than 5-years from index date, for various registries, causing variable impact on survival estimate. When differential losses to follow up were further detected with respect to potential determinants, either of losses or survival, it became necessary to make specific loss-adjustment for correcting the ensuing bias otherwise estimated by standard methods. Large differences between loss-adjusted and actuarial survival from female breast cancer were found in hospital-based series of patients coming from wide geographic area, where follow up of patients, no longer attending the hospital, by house visits was impractical (Ganesh, 1995). However, in another study from Trivandrum hospital study of ovarian cancer, the differences between loss-adjusted survival before active follow up and survival by life table approach after perseverant postal enquiries were small (Mathew, 1996). The response was directly dependent on the effort expended by individual registries to collect vital status information. In contrast, an international comparison of actuarial and loss adjusted survival of cervix cancer incident cases from population-based registries in developing countries (Swaminathan et al., 2002) found a maximum difference of 4.1% with a loss to follow up to 44% but indicating practically a random pattern. The difference was 2.1 units in Khon Kaen registry in Thailand for cervix cancer. These minimal differences are mainly due to the integration of mortality data collection into the case-finding operations of any population-based registry, routinely done on an annual basis and their matching with incident cancer database. This is carried out for all cases registered irrespective of any selection. This induces

the randomness in cases on whom, the vital status is unknown. Loss adjusted survival was lesser than actuarial one, which suggests that under ascertainment of deaths among loss to follow up cases is the problem. The only exception were patients with localized or early stage breast cancer wherein, the loss-adjusted survival exceeded the actuarial survival indicating that non-tracing of alive patients till closing date was the problem (Ganesh, 1995). Hence, data source seems to affect the need for loss-adjustment and the problem can be more substantial in hospital-based cancer registries and clinical series. The loss-adjusted approach is likely to be useful especially when hospital-based cancer registry data from a low or medium resource country are used to evaluate the outcome of cancer patients. If routine follow up is poor, first priority is to increase the actual follow up visits and the second is to institute rigorous active follow up measures for eliciting the vital status.

6.4 Intra-country variation in cancer survival – Methodological implications

Comparing the maximum and minimum survival by registries in the same country in Table 17, there were marked variations for all cancers in China, most cancers in India, a few cancers in Thailand and laryngeal cancer in South Korea. In India, marked differences were observed between rural (Barshi) or semi-urban (Karunagappally) or small urban (Bhopal) registries while the differences were minimal for most cancers between the metropolitan cities of Chennai and Mumbai, where more developed and accessible cancer care services are available. The follow up data was obtained entirely by active methods in all registries from India and hence the differences in survival due to methodological bias are likely to be minimal. In South Korea, the survival differences between registries were negligible for all cancers excepting larynx. Follow up was carried out predominantly by passive follow up methods, primarily by data linkages using unique national identity numbers issued to citizens and scrutiny of hospital records. Hence, the mismatches are likely to be lesser and so would be the methodological bias. However, the registries from China and Thailand have adopted

different follow up methods. In China, cancer survival from rural area of Qidong was consistently lower than from big cities like Shanghai and Tianjin for all cancers. In Thailand, the differences between registries were limited to a few cancers only. These survival differences are likely to reflect the differences in early diagnosis and the impact of inequities in health care infrastructure development, availability and accessibility to health services, but the differences due to bias in estimating survival following methodological problems are inherent but unknown. It is encouraged that future studies address this important issue and facilitate objective measurement of possible bias in the estimated survival owing to inappropriate choice of methods of follow up for vital status information.

6.5 Survival differences – Implications of treatment or disease characteristics

The success of cancer treatment is, as a rule, measured by survival. The variable level of development of cancer health services certainly impacts the survival from different populations in low or medium resource countries calling for adequate and sincere investments in improving awareness, health-services infrastructure and accessibility. Cancer survival from both hospital or population series have different perspectives but serve as the main indicator of outcome of cancer health services or treatment and an important component in maintaining cancer control activities. Reliable statistics devoid of methodological bias in eliciting vital status are required from low or medium resource countries for specifically interpreting any survival differences in the region or institution as due to treatment related attributes or resources.

Table 19: Comparison of 5-year observed survival (%) of major cancers in rural Dindigul registry^a, urban Chennai registry (II) and treated cases from hospital cancer registry in Cancer Institute (WIA), Chennai^c

Cancer site/type	Five-year observed survival %		
	Dindigul rural PBCR	Chennai urban PBCR	Treated cases only from HBCR
	2003	1990-99	2000-01
Cervix	35	54	62
Breast	38	44	67
Stomach	7	9	41
Mouth	33	31	35
Oesophagus	6	7	14
Tongue	27	19	37
Lung	4	7	8
All leukaemias	19	20	44
All lymphomas	25	29	61
Larynx	25	31	70
Ovary	30	27	45
Large bowel	16	NP	42
All cancers	26	NP	NP

PBCR: Population based cancer registry – comprises all incident cases treated or not
HBCR: Hospital based cancer registry – comprises all cases that completed one treatment
modality at the Cancer Institute (WIA), Chennai, India; NP: Not published;

^a Swaminathan et al., (2009); ^c Shanta et al., (2008)

Table 19 gives the comparison of survival estimates arising from rural (Dindigul, India) and urban (Chennai, India) population-based registries and hospital series of treated cases (Cancer Institute (WIA), Chennai, India) of various cancers, from the same state of Tamil Nadu in South India. Cancer follow-up places a significant burden on hospital outpatient clinics and hence alternative models need to be developed to provide the same.

The follow up for vital status information on patients was carried out entirely by active methods in all three registries. The data quality indices like losses to follow up have all been addressed and were uniform in the three registries. Hence, survival differences, if any, could be attributed to non-methodological factors like treatment or disease characteristics. Hospital-based survival studies generally suffer from high degree of selection of patients with favourable prognosis and hence usually lack representativeness, thereby not always suitable for generalization to a larger population in the region. However, analysis of overall or event-free survival need not be a sterile exercise if it contrives on improving clinical practice, which would translate into ultimate benefit at the population level. Population-based survival in rural Dindigul district (Swaminathan et al., 2009) characterized by all cases treated or not was either on par or lower than in metropolitan Chennai and other urban registries in India (Sankaranarayanan and Swaminathan, 2011). Small differences in survival between rural and metropolitan Chennai registries are not reflective of wide differentials in availability or development of or accessibility to cancer related health services in these two places. This brings to the fore, the necessity of collecting basic data on prognostic factors (like extent of disease, treatment, etc.) even in a population based series of cancer cases to explain the survival differences meaningfully. This extra data would also help to understand and enhance the data quality issues on survival estimation.

Though cancer treatment, including chemotherapy, is given free of cost in all public or government hospitals in Tamil Nadu state, patient compliance and completeness of treatment are significant prognostic factors for survival. This view is strengthened since survival witnessed among all patients receiving complete treatment in Cancer Institute (WIA), Chennai, a comprehensive cancer centre with state of art facilities, was two-fold higher than in rural Dindigul district and was either on par or higher than population-based survival in Europe for most cancers (Sant et al., 2009, Swaminathan et al., 2009). While the

survival in public hospitals was less than in non-government sector for patients who received complete treatment, the reverse was true for incomplete treatment, as demonstrated in the present study on population series of childhood cancers from Chennai. This probably reflects the impact of stage of disease, socioeconomic status, type of treatment received and compliance and supportive care on survival. Unlike in other developed countries, it is possible that economic constraints in living conditions may affect completeness of treatment more than cost of treatment itself at the population level in rural or urban India. Commissioning of special cancer registries as an extension of HBCRs to address the variations in absolute survival is one thoughtful solution. This would ensure data collection on a variety of important treatment and disease characteristics. Population based trials focusing on technologically and economically viable early detection programs for major cancers allied to accessible treatment facilities are the way forward to improve cancer outcome.

Tumour stage is one of the important disease characteristics that impacts survival and provides the basis for differences in survival. Information on clinical extent of disease was available for selected cancers in a few population-based registries from low or medium resource countries. This was utilized to examine and compare the survival differences by stage between registries classified into two groups: possessing well (W) or moderately (M) developed cancer health services based on per head GNI values. Since data collection was from heterogeneous sources, misclassification in extent of disease categories is a distinct possibility. To minimize this, the data from individual registries in respective groups were pooled and analyzed for comparison (Table 20).

Table 20: 5-year absolute survival% and frequency% distribution by clinical extent of disease of selected cancers based on two sets of pooled data from contributing countries (I)

Well developed health services (W) vs. moderately developed (M): Cancer site and countries (Pooled number of cases)	5-year survival% by clinical extent of disease (Frequency% by clinical extent of disease)			
	Localised	Regional	Distant Metastasis	Unknown
<u>Tongue</u>				
W. Survival: Singapore (120)	48.4	23.3	20.0	33.1
Frequency	(26)	(25)	(4)	(45)
M. Survival: India, Pakistan, Thailand (3,844)	54.3	14.5	3.1	25.3
Frequency	(23)	(65)	(5)	(7)
<u>Oral cavity</u>				
W. Survival: Singapore (135)	52.3	26.5	-	28.9
Frequency	(28)	(22)	(2)	(48)
M. Survival: India, Pakistan, Thailand (5,592)	60.2	23.8	3.3	28.8
Frequency	(22)	(66)	(5)	(7)
<u>Large bowel</u>				
W. Survival: Singapore, Turkey (4,969)	64.1	45.7	8.6	41.8
Frequency	(26)	(27)	(18)	(29)
M. Survival: India, Philippines, Thailand (4,742)	49.8	32.0	2.4	34.7
Frequency	(29)	(34)	(23)	(14)
<u>Larynx</u>				
W. Survival: Singapore, Turkey (789)	69.6	40.7	41.8	54.6
Frequency	(34)	(19)	(5)	(42)
M. Survival: India, Thailand (3,161)	54.4	22.3	4.7	26.9
Frequency	(25)	(60)	(8)	(7)
<u>Breast</u>				
W. Survival: China (Hong Kong), Singapore, Turkey (14,645)	89.6	75.4	26.7	79.7
Frequency	(17)	(32)	(2)	(49)
M. Survival: Costa Rica, India, Philippines, Saudi Arabia, Thailand (17,640)	76.3	47.4	14.9	47.1
Frequency	(26)	(47)	(14)	(13)
<u>Cervix</u>				
W. Survival: Singapore, Turkey (1,230)	69.5	52.2	18.6	57.5
Frequency	(42)	(13)	(5)	(40)
M. Survival: Costa Rica, India, Philippines, Thailand (14,536)	73.2	47.2	7.4	45.7
Frequency	(20)	(64)	(6)	(10)
<u>Ovary</u>				
W. Survival: Singapore, Turkey (948)	84.1	39.7	28.1	56.7
Frequency	(40)	(4)	(27)	(29)
M. Survival: India, Thailand (3,666)	63.8	34.5	4.2	36.8
Frequency	(22)	(27)	(38)	(13)
<u>Bladder</u>				
W. Survival: Singapore, Turkey (1,062)	61.3	34.8	16.4	54.0
Frequency	(53)	(7)	(5)	(35)
M. Survival: India, Thailand (2,476)	43.8	24.9	2.3	35.6
Frequency	(42)	(33)	(10)	(15)

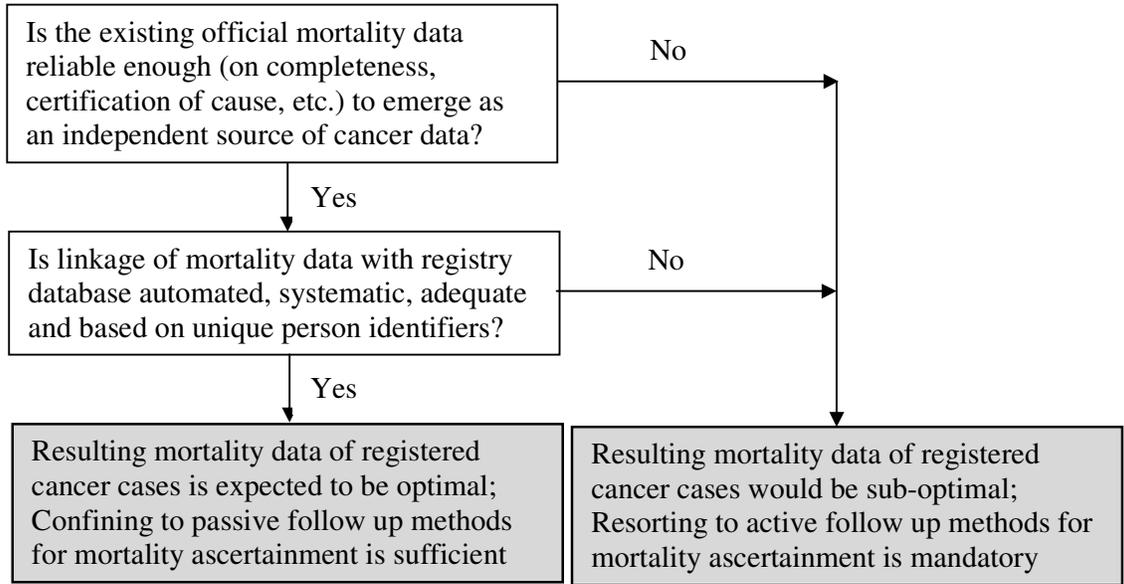
There was an inverse relationship between stage of disease or extent of disease and survival in both groups for all cancers which added strength to data quality (Table 20). The pattern of survival by clinical extent of disease in the study provides striking evidence for the need for early diagnosis and effective treatment. The higher survival observed for group W countries for localized large bowel, larynx, breast, ovary and bladder cancers (which largely require radical surgical treatment) and for regional spread diseases (which require multimodal treatment) may largely due to the difference in the development and accessibility of diagnostic and treatment services.

Cancer survival studies from low or medium resource countries would not have been possible without the availability of reliable population-based cancer registries. Such studies form the first step towards indentifying the elements of cancer control, including primary prevention, early detection initiatives and treatment, which are most likely to contribute to the reduction of cancer mortality in these countries. These studies are also pointers towards realization that cancer survival in many low-income or middle-income countries that are yet to undertake such studies is likely to be on par or lower than discussed here. The survival rates observed in many countries have also provided a definite alternative to the deficient official cancer mortality statistics from those areas. Extreme caution must be exercised when survival figures from local area registries are used to extrapolate for the entire country. For instance, in India, the coverage of PBCRs is about 7% of national population. The PBCRs are predominantly covering metropolitan or urban populations. With few registries covering rural populations, even pooled data analysis of all registries would result in erroneous estimates, mostly representing experiences in urban areas only. However, the striking differences in cancer survival between countries emphasize the need for urgent and adequate investments on cancer control.

It must be kept in mind that in a PBCR, using data from specialized and non-specialized medical institutions of varying standards (heterogeneous) could lead to unrealistic comparisons between registries. However, methodological issues on estimating survival, including follow up data acquisition, are entirely within the control of individual registries. A simple guide to PBCRs for getting vital status data of registered cases in the conduct of survival studies is given in figure 4.

It cannot be reiterated more that cancer survival studies based on registry data from low or medium resource countries do need suitable correction for possible inherent methodological bias, arising due to inappropriate employment of follow up methods (like relying on passive methods under sub-optimal mortality registration) and/or non-specific methods for estimation of survival (like standard actuarial instead of loss-adjusted ones). It can then stated with overwhelming confidence that these survival data generated from low or medium resource countries provided a baseline for sincere investments in developing infrastructure for sustainable improvements in cancer health services in the future.

I. Ascertainment of mortality of cancer cases from official vital statistics offices



II. Ascertainment of vital status of presumably alive or all registered cancer cases

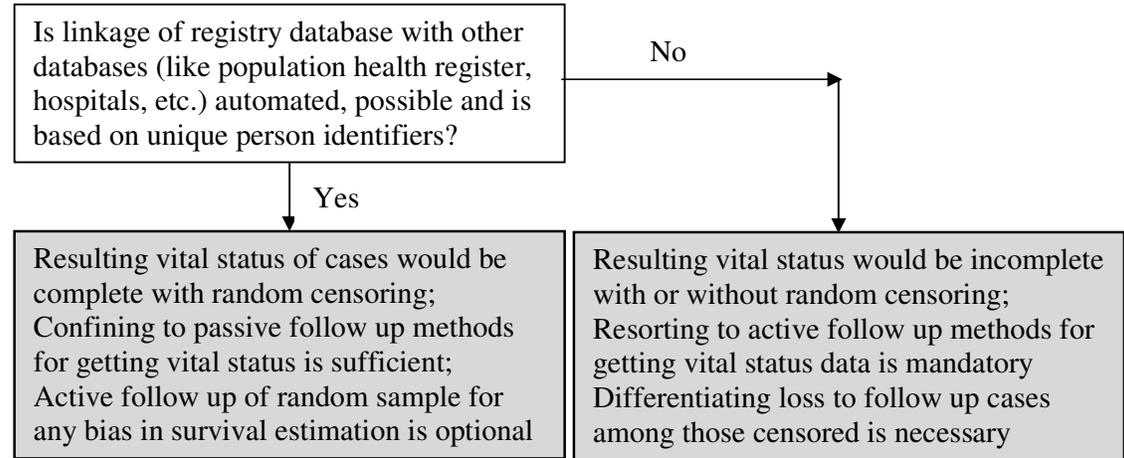


Figure 4: A guide for choosing follow up methods for conducting population-based survival study in low or medium resource settings

7. CONCLUSIONS

- The execution of survival studies in low or medium resource countries setting requires special efforts and resources in terms of personnel, expertise and funding. Unlike in more developed countries, survival studies could not always be routinely carried out, given the less developed health information systems.
- In less developed health information systems, mortality ascertainment by passive means would be grossly inadequate or incomplete, inducing a serious bias in survival estimation, if standard vital status assumptions were followed (like cases always presumed to be alive until receiving death notification).
- To avoid this upward bias in survival estimation, the maximum ranging between 22-47% for different cancers as shown in this study, a variety of suitable active methods have to be evolved and pursued for collecting vital status information.
- If active methods are impractical to implement owing to registry operational constraints, active follow up of representative subset of cases should be systematically undertaken to elucidate the bias in estimated survival under different assumptions on the vital status as done in this study.
- If high magnitude of non-random loss to follow up exists, its determinants have to be elicited and survival estimation done through differential loss-adjustment procedures. The impact could be variable: minimal for population-based but would be pronounced for hospital-based studies as shown in this study.

- A systematic evaluation of biases in estimating survival due to methodological problems and their suitable corrections are mandatory before survival differences could be attributed to the varied development of treatment resources and/or disease characteristics in low or medium resources settings.

8. ACKNOWLEDGMENTS

Words are inadequate to express my sincere thanks and profound gratitude to my research supervisor and guide, Dr. M. Hakama, Emeritus Professor, School of Health Sciences, (formerly Tampere School of Public Health) University of Tampere. His unstinted cooperation by sparing unlimited time to evaluate the original manuscripts and his valuable inputs to enhance the quality of this work are unmatched. I am indebted to him, for his invaluable guidance, immense patience and all the support.

I owe my position entirely to Dr. V. Shanta, Chairman, Cancer Institute (W.I.A), Chennai. She has been a constant source of motivation and encouragement to me on all the aspects for pursuing my doctoral research in epidemiology. I am thankful to her for all the support.

The International Postgraduate Programme in Epidemiology (IPPE) is a boon to students from international community wanting to excel in epidemiology and I thank the University of Tampere for conceiving this programme and for the total funding.

I gratefully recall the continued opportunity provided to me by Dr. R. Sankaranarayanan, Head, Screening Group, International Agency for Research on Cancer, Lyon, France, to work with multinational data on cancer survival since 1995. I thank him for giving the permission to utilize the required data from SURVCAN databases.

I thank the official referees, Dr. C. Varghese, Technical Officer, Western Pacific Regional Office of WHO, Manila, Philippines, and Dr. C. Ramesh, Professor and Head, Kidwai Memorial Institute of Oncology, Bangalore, India, for consenting to review the dissertation and for providing critical comments and valuable inputs to improve the manuscript.

My sincere thanks to my teachers, Dr. A. Auvinen, Dr. S. Virtanen, Ms. H. Huhtala, Dr. T. Salminen, from Tampere School of Public Health; Dr. T. Hakulinen from Finnish Cancer Registry; Dr. N. Fieller from UK and many others. I appreciate the help of Ms. Catarina Stahle-Nieminen, International Coordinator, for providing a hassle-free environment and all assistance right from the day of first landing at Tampere and during subsequent visits.

I shall be failing in my duty if I do not recognize the wholehearted contribution made by all the co-authors of the original publications that formed the basis of this dissertation and other collaborators of the SURVCAN project.

I fondly recall the times that I spent with my fellow students of IPPE course from India, Dr. C. Varghese, Dr. S. Jayadevan, Dr. A. Budukh, Dr. M. Dhar and Mr. J. Jayaram, for making my stay in Tampere, a thoroughly enjoyable experience in my life.

I am extremely grateful to all my colleagues in the department of Biostatistics and Cancer Registry, Cancer Institute (WIA), Chennai, for their diligent work and support.

I firmly believe that my father, who is no more, has been the guiding spirit in my endeavour. I owe my gratitude to my wife, son, mother and in-laws, for their fullest cooperation and support.

9. REFERENCES

- Aareleid T, Gondos A, Brenner H, Pokker H, Leppik K and Mägi M (2011): Testicular cancer survival in Estonia: improving but still relatively low. *Acta Oncol* 50(1):99-105.
- Berg JW, Ross R and Latourette HB (1977): Economic status and survival of cancer patients. *Cancer* 39:467-77.
- Berrino F, Sant M, Verdecchia A, Capocaccia R, Hakulinen T and Esteve J (1995): Survival of Cancer Patients in Europe: the EURO CARE Study. IARC Scientific Publications No. 132. International Agency for Research on Cancer, Lyon.
- Black RJ and Bashir SA (1998a): World standard cancer patient populations: A resource for comparative analysis of survival data. In: *Cancer Survival in Developing Countries*, pp 9-11. Eds. R. Sankaranarayanan, RJ Black and DM Parkin. IARC Scientific Publications No. 145, International Agency for Research on Cancer, Lyon.
- Black RJ and Swaminathan R (1998b): Statistical methods for the analysis of cancer survival data. In: *Cancer Survival in Developing Countries*, pp 3-7. Eds R. Sankaranarayanan, RJ Black and DM Parkin. IARC Scientific Publications No. 145, International Agency for Research on Cancer, Lyon.
- Black RJ, Sankaranarayanan R and Parkin DM (1998c): Interpretation of population based cancer survival data. In: *Cancer Survival in Developing Countries*, pp 13-17. Eds. R. Sankaranarayanan, RJ Black and DM Parkin. IARC Scientific Publications No. 145, International Agency for Research on Cancer, Lyon.
- Brenner H, Gefeller O and Hakulinen T (2002): A computer program for period analysis of cancer patient survival. *Eur J Cancer* 38:690-95.
- Brenner H and Gefeller O (2004a). An alternative approach to monitoring cancer patient survival. *Cancer* 78:2004-10.

- Brenner H, Arndt V, Gefeller O and Hakulinen T (2004b). An alternative approach to age adjustment of cancer survival rates. *Eur J Cancer* 40:2317–22.
- Brenner H and Hakulinen T (2009): Implications of incomplete registration of deaths on long-term survival estimates from population-based cancer registries. *Int J Cancer* 125:432-37.
- Cavalli-Björkman N, Lambe M, Eaker S, Sandin F, Glimelius B (2011): Differences according to educational level in the management and survival of colorectal cancer in Sweden. *Eur J Cancer* 47(9):1398-406.
- Chia KS (2011): Cancer survival in Singapore, 1993-1997. In: *Cancer survival in Africa, Asia, the Caribbean and Central America*, pp 183-98. Eds. R Sankaranarayanan and R Swaminathan, IARC Scientific Publications No. 162, International Agency for Research on Cancer, Lyon.
- Chiang CL (1968): *Introduction to stochastic processes in Biostatistics*. Wiley Publications, New York.
- Coleman MP, Quaresma M, Berrino F, Lutz JM, De Angelis R, Capocaccia R, Baili P, Rachet B, Gatta G, Hakulinen T, Micheli A, Sant M, Weir HK, Elwood JM, Tsukuma H, Koifman S, E Silva GA, Francisci S, Santaquilani M, Verdecchia A, Storm HH, Young JL and CONCORD Working Group (2008): Cancer survival in five continents: a worldwide population-based study (CONCORD). *Lancet Oncol* 9:730–56.
- Cox DR (1972): Regression models and life tables. *J R Statistic Soc (B)* 34:187-220.
- Curado MP, Edwards B, Shin HR, Storm H., Ferlay J, Heanue M and Boyle P (2007): *Cancer Incidence in Five Continents, Vol. IX*. IARC Scientific Publications No. 160. International Agency for Research on Cancer, Lyon.
- Cutler SJ and Ederer F (1958): Maximum utilization of the life table method in analysing survival. *J Chron Dis* 8:699–712.
- Ederer F, Axtell LM and Cutler SJ (1961): *The relative survival rate: a statistical methodology*. *Monogr Natl Cancer Inst* 6:101–21.

Eser S (2011): Cancer survival in Izmir, Turkey, 1995-1997. In: Cancer survival in Africa, Asia, the Caribbean and Central America, pp 237-42. Eds. R Sankaranarayanan and R Swaminathan, IARC Scientific Publications No. 162, International Agency for Research on Cancer, Lyon.

Feinstein AR, Sosin DM and Wells CK (1985): The Will Rogers phenomenon: stage migration and new diagnostic techniques as a source of misleading statistics for survival in cancer. *New Engl J Med* 312:1604-08.

Ferlay J, Bray F, Pisani P and Parkin DM (2004): GLOBOCAN 2002: Cancer incidence, mortality and prevalence worldwide. IARC CancerBase No. 5, version 2.0, IARC Press, Lyon.

<http://www-dep.iarc.fr/> (accessed in March 2009)

Ganesh B (1995): Effect of loss to follow up in estimating survival rates. *Acta Universitatis Tamperensis, Ser A, Vol 440*. University of Tampere, Tampere.

Garrote LF, Alvarez YG, Babie PT, Yi MG, Alvarez MG and Cicili ML (2011): Cancer survival in Cuba, 1994-1995. In: Cancer survival in Africa, Asia, the Caribbean and Central America, pp 89-95. Eds. R Sankaranarayanan and R Swaminathan, IARC Scientific Publications No. 162, International Agency for Research on Cancer, Lyon.

Hakulinen T (1982): Cancer survival corrected for heterogeneity in patient withdrawal. *Biometrics* 38:933-42.

Hanai A and Fujimoto I (1985): Survival as an index in evaluating cancer control. In: The role of the registry in cancer control. Eds. D.M. Parkin, G. Wagner and C.S. Muir. IARC Scientific Publications No. 66, International Agency for Research on Cancer, Lyon.

Hermanek P and Sobin LH (1987): TNM classification of malignant tumours, fourth fully revised edition. Springer-Verlag, London.

Howlader N, Noone AM, Krapcho M, Neyman N, Aminou R, Waldron W, Altekruse SF, Kosary CL, Ruhl J, Tatalovich Z, Cho H, Mariotto A, Eisner MP, Lewis DR, Chen HS, Feuer EJ,

- Cronin KA and Edwards BK (2011). SEER Cancer Statistics Review, 1975-2008. National Cancer Institute, Bethesda MD.
- ICD-10 International Statistical Classification of Diseases and Related Health Problems (1992): Tenth revision, Volume 1. World Health Organization: Geneva.
- ICD-O International Classification of Diseases for Oncology (2000): Third Edition. World Health Organization: Geneva.
- Jensen OM, Parkin DM, Maclellan R, Muir CS and Skeet RG (1991): Cancer Registration Principles and Methods. IARC Scientific Publications No. 95. International Agency for Research on Cancer, Lyon.
- Kaplan EL and Meier P (1958): Non parametric estimation from incomplete observation. J Amer Stat Assoc 53:457-81.
- Karjalainen S and Palva I (1989): Do treatment protocols improve end results? A study of survival of patients with multiple myeloma in Finland. BMJ 299(6707):1069-72.
- Kogevinas M (1991): Longitudinal study: Socio-demographic differences in cancer survival. OPCS Series L.S. No. 5, HMSO: London.
- Krishnamurthi S, Shanta V and Sastri DV (1971): Combined therapy in buccal mucosal cancers. Radiology 99(2):409-15.
- Laudico AV, Esteban D and Parkin DM (1989). Cancer in the Philippines. IARC Technical Report No. 5, International Agency for Research on Cancer, Lyon.
- Lopez AD, Ahmad OB, Guillot M, Inoue M, Ferguson BD and Salomon JA (2001): Life tables for 191 countries for 2000: Data, Methods, Results (GPE discussion paper No. 40). Health Systems Performance Assessments Peer Review, Technical Documentation IV Outcomes: Population Health. Evidence and Information for Policy (EIP). World Health Organization, Geneva.

- Magrath I and Litvak J (1993): Cancer in developing countries: opportunity and challenge. *J Natl Cancer Inst* 85:862–74.
- Mathew A (1996): Removing bias in cancer survival estimates by active follow up and information on determinants of loss to follow up. *Acta Universitatis Tamperensis, Ser A, vol 525*. University of Tampere, Tampere.
- Mervic L, Leiter U, Meier F, Eigentler T, Forschner A, Metzler G, Bartenjev I, Büttner P and Garbe C (2011): Sex differences in survival of cutaneous melanoma are age dependent: an analysis of 7338 patients. *Melanoma Res.* 21(3):244-52.
- Muir CS, Demaret E and Boyle P (1985): The cancer registry in cancer control: an overview. In: *The role of the registry in cancer control*. IARC Scientific Publications No. 66. Eds. DM Parkin, G Wagner and CS Muir, International Agency for Research on Cancer, Lyon.
- Nair MK, Sankaranarayanan R, Nair KS, Amma NS, Varghese C, Padmakumari G and Cherian T (1993): Overall survival from breast cancer in Kerala, India, in relation to menstrual, reproductive and clinical factors. *Cancer* 71(5):1791-6.
- Nandakumar A, Anantha N, Venugopal TC, Sankaranarayanan R, Thimmasetty K, Dhar M (1995): Survival in breast cancer: a population-based study in Bangalore, India. *Int J Cancer* 60(5):593-6.
- National Cancer Registry Programme (1992): Biennial Report: 1988-89. Indian Council of Medical Research, New Delhi.
- Parkin DM, Ferlay J, Hamdi-Chérif M, Sitas F, Thomas JO, Wabinga H and Whelan SL (2003): *Cancer in Africa*. IARC Scientific Publications No. 153. International Agency for Research on Cancer, Lyon.
- Parkin DM, Whelan SL, Ferlay J and Storm H (2005): *Cancer incidence in five continents, Vol I to VIII*. IARC Cancerbase No. 7. International Agency for Research on Cancer, Lyon.

- Pavlovsky S, Sackmann MF, Santarelli MT, Svarch E, Jimenez E, Kohan R and Rosso A (1992):
An update of the results of intensive therapy in children with acute lymphoblastic leukaemia.
Leukemia 6 (Suppl 2):167-70.
- Pengsaa P, Pesi M, Udomthavornsuk B, Tungvorapongchai V, Vatanasapt W and Shibata Y
(1989): Survival analysis of cervical cancer patients in Srinagaraind hospital 1976-1987. J
Med Assoc Thai 72(6):346-50.
- Pokhrel A, Martikainen P, Pukkala E, Rautalahti M, Seppä K, Hakulinen T (2010): Education,
survival and avoidable deaths in cancer patients in Finland. Br J Cancer 103(7):1109-14.
- Rao DN, Shroff PD, Chattopadhyay G and Dinshaw KA (1998). Survival analysis of 5595 head
and neck cancers-results of conventional treatment in a high risk population. Br J Cancer
77:1514-1518.
- Redaniel MT, Laudico A, Mirasol-Lumague MR, Gondos A, Pulte D, Mapua C, Brenner H (2009):
Cancer survival discrepancies in developed and developing countries: comparisons between
the Philippines and the United States. Br J Cancer 100(5):858-62.
- Ries LAG, Harkins D, Krapcho M, Mariotto A, Miller BA, Feuer EJ, Clegg L, Eisner MP, Horner
MJ, Howlander N, Hayat M, Hankey BF, Edwards BK (2006): SEER Cancer Statistics
Review, 1975-2003, National Cancer Institute, Bethesda MD.
- Roberts JC, Li G, Reitzel LR, Wei Q, Sturgis EM (2010): No evidence of sex-related survival
disparities among head and neck cancer patients receiving similar multidisciplinary care: a
matched-pair analysis. Clin Cancer Res. 16(20):5019-27.
- Sankaranarayanan R, Black RJ and Parkin DM (1998): Cancer Survival in Developing Countries.
IARC Scientific Publications No. 145. International Agency for Research on Cancer, Lyon.
- Sankaranarayanan R and Swaminathan R (2011): Cancer survival in Africa, Asia, the Caribbean
and Central America, IARC Scientific Publications No. 162. International Agency for
Research on Cancer, Lyon.

- Sant M, Allemani C, Santaquilani M, Knijn A, Marchesi F, Cappocaccia R and the EURO CARE Working Group (2009) : EURO CARE-4. Survival of cancer patients diagnosed in 1995-1999. Results and commentary. *Eur J Cancer* 45:931–91.
- Shanta V, Gajalakshmi CK, Swaminathan R, Ravichandran K and Vasanthi L (1994): Cancer registration in Madras Metropolitan Tumour Registry, India. *Eur J Cancer* 30:974-78.
- Shanta V, Swaminathan R and Rama R (2008): Biennial Report 2005-2006. Hospital Cancer Registry. Cancer Institute (WIA), Chennai.
- Shi Q, You YN, Nelson H, Allen MS, Winchester D, Stewart A, Young-Fadok T, Decker PA, Green EM, Holton SJ, Ballman KV (2010): Cancer registries: a novel alternative to long-term clinical trial follow-up based on results of a comparative study. *Clin Trials* 7(6):686-95.
- Sierra R, Barrantes R, Munoz-Leiva G, Bieber CA and Munoz-Calero A (1988): Cancer in Costa Rica. IARC Technical Report No. 1, International Agency for Research on Cancer, Lyon.
- Sriamporn S, Black RJ, Sankaranarayanan R, Kamsa Ad S, Parkin DM and Vatanasapt V (1995): Cancer survival in Khon Kaen province, Thailand. *Int J Cancer* 61:296-300.
- Sriplung H and Prechavittayakul P (2011): Cancer survival in Songkhla, Thailand, 1990-1999. In: Cancer survival in Africa, Asia, the Caribbean and Central America, pp 227-35. Eds. R Sankaranarayanan and R Swaminathan, IARC Scientific Publications No. 162, International Agency for Research on Cancer, Lyon.
- Stiller CA (1994): Population based survival rates for childhood cancer in Britain, 1980-91. *BMJ* 309(6969):1612-6.
- SURVCAN database (2011): Cancer survival in Africa, Asia, the Caribbean and Central America. Eds. R. Sankaranarayanan, R. Swaminathan and E. Lucas. International Agency for Research on Cancer: Lyon. (<http://survcan.iarc.fr/survival.php>) (accessed on May 10, 2011)
- Suwanrungruang K, Vatanasapt P, Kamsa-Ard S, Sriamporn S and Wiangnon S (2011): Cancer survival in Khon Kaen, Thailand, 1993-1997. In: Cancer survival in Africa, Asia, the

- Caribbean and Central America, pp 211-16. Eds. R Sankaranarayanan and R Swaminathan, IARC Scientific Publications No. 162, International Agency for Research on Cancer, Lyon.
- Swaminathan R, Black RJ and Sankaranarayanan R (1998). Database on Cancer Survival from Developing Countries. In: Cancer survival in developing countries, pp 19-25. Eds. R Sankaranarayanan, RJ Black and DM Parkin, IARC Scientific Publications No. 145, International Agency for Research on Cancer, Lyon.
- Swaminathan R, Sankaranarayanan R, Hakama M and Shanta V (2002): Effect of loss to follow up on population based cancer survival rates in developing countries. *Int J Cancer Suppl.* 13:172.
- Swaminathan R, Selvakumaran R, Esmey PO, Sampath P, Ferlay J, Jissa V, Shanta V, Cherian M and Sankaranarayanan R (2009): Cancer pattern and survival in a rural district in South India. *Cancer Epidemiol* 33(5):325-31.
- Swaminathan R, Lucas E and Sankaranarayanan R (2011): Cancer survival in Africa, Asia, the Caribbean and Central America: Database and attributes. In: Cancer survival in Africa, Asia, the Caribbean and Central America, pp 23-31. Eds. R Sankaranarayanan and R Swaminathan, IARC Scientific Publications No. 162, International Agency for Research on Cancer, Lyon.
- Valsecchi MG and Steliarova-Foucher E (2008): Cancer registration in developing countries: luxury or necessity? *Lancet Oncol* 9(2):159-67.
- Vatanasapt V, Martin N, Sriplung H, Chindavijak K, Sontipong S, Sriamporn S, Parkin DM and Ferlay J (1993). Cancer in Thailand 1988-1991. IARC Technical Report No. 16. International Agency for Research on Cancer, Lyon.
- Young JL (1991). The hospital-based cancer registry. In: Cancer registration principles and methods, IARC Scientific Publications No. 95, pp 177-184. Eds. OM Jensen, DM Parkin, R MacLennan, CS Muir and RG Skeet. International Agency for Research on Cancer, Lyon.

ORIGINAL PUBLICATIONS

Cancer survival in Africa, Asia, and Central America: a population-based study



Rengaswamy Sankaranarayanan, Rajaraman Swaminathan, Hermann Brenner, Kexin Chen, Kee Seng Chia, Jian Guo Chen, Stephen C K Law, Yoon-Ok Ahn, Yong Bing Xiang, Balakrishna B Yeole, Hai Rim Shin, Viswanathan Shanta, Ze Hong Woo, Nimit Martin, Yupa Sumitsawan, Hutcha Sriplung, Adolfo Ortiz Barboza, Sultan Eser, Bhagwan M Nene, Krittika Suwanrungruang, Padmavathamma Jayalekshmi, Rajesh Dikshit, Henry Wabinga, Divina B Esteban, Adriano Laudico, Yasmin Bhurgri, Ebrima Bah, Nasser Al-Hamdan

Summary

Background Population-based cancer survival data, a key indicator for monitoring progress against cancer, are not widely available from countries in Africa, Asia, and Central America. The aim of this study is to describe and discuss cancer survival in these regions.

Methods Survival analysis was done for 341 658 patients diagnosed with various cancers from 1990 to 2001 and followed up to 2003, from 25 population-based cancer registries in 12 countries in sub-Saharan Africa (The Gambia, Uganda), Central America (Costa Rica), and Asia (China, India, Pakistan, Philippines, Saudi Arabia, Singapore, South Korea, Thailand, Turkey). 5-year age-standardised relative survival (ASRS) and observed survival by clinical extent of disease were determined.

Findings For cancers in which prognosis depends on stage at diagnosis, survival was highest in China, South Korea, Singapore, and Turkey and lowest in Uganda and The Gambia. 5-year ASRS ranged from 76–82% for breast cancer, 63–79% for cervical cancer, 71–78% for bladder cancer, and 44–60% for large-bowel cancers in China, Singapore, South Korea, and Turkey. Survival did not exceed 22% for any cancer site in The Gambia; in Uganda, survival did not exceed 13% for any cancer site except breast (46%). Variations in survival correlated with early detection initiatives and level of development of health services.

Interpretation The wide variation in cancer survival between regions emphasises the need for urgent investments in improving awareness, population-based cancer registration, early detection programmes, health-services infrastructure, and human resources.

Funding Association for International Cancer Research (AICR; St Andrews, UK), Association pour la Recherche sur le Cancer (ARC, Villejuif, France), and the Bill & Melinda Gates Foundation (Seattle, USA).

Introduction

Cancer survival estimates from population-based cancer registries include all cases diagnosed in a given population. These estimates reflect different socioeconomic factors, health-care seeking behaviours, natural histories, and the efficiency of the health-care services to provide early diagnosis, prompt treatment, and follow-up care. Population-based survival represents the average prognosis of a cancer and is useful for assessing progress in cancer control, including the effect of early detection, diagnosis, treatment, and follow-up on cancer outcomes. These data are also helpful in making informed decisions to ensure improved and equitable cancer care. The International Agency for Research on Cancer (IARC) has been collating data on worldwide cancer incidence for five decades,^{1,2} in collaboration with the International Association of Cancer Registries and registries in various countries, with a particular focus on low-income (per head Gross National Income [GNI] <US\$2000) and middle-income countries (per head GNI US\$2000–10 000). Such efforts have been complemented by WHO mortality databases and population-based survival studies that systematically analysed survival outcomes of patients in

Europe,³ the USA,⁴ and other developed countries.^{5–7} Cancer survival statistics from ten developing countries were made available for the first time, to our knowledge, in 1995 through a collaborative initiative by IARC,^{8,9} but such data are not widely available from many countries of low to middle income. Here, we report the results from a collaborative survival study by IARC with a wider geographical coverage of countries and populations.

Methods

Registries

31 population-based cancer registries in 17 countries provided data for this study; six population-based cancer registries from five countries were excluded from participation. The methods used in each geographical region to identify and register all diagnosed cases are described in the technical reports from the individual registries and in the Cancer Incidence in Five Continents (CI5) series published by the IARC.^{1,2} The registries used a mix of passive notification and active registration to register all cancer cases diagnosed in their area, and used quality assurance procedures, as advocated by IARC, to validate the quality and completion of cancer registration

Lancet Oncol 2010; 11: 165–73

Published Online
December 10, 2009
DOI:10.1016/S1470-
2045(09)70335-3

See [Reflection and Reaction](#)
page 110

Screening Group, International Agency for Research on Cancer, Lyon, France
(R Sankaranarayanan MD, H R Shin MD); Division of Epidemiology and Cancer Registry, Cancer Institute (WIA), Chennai, India (R Swaminathan PhD, V Shanta DSc); Division of Clinical Epidemiology and Aging Research, German Cancer Research Center, Heidelberg, Germany (H Brenner MD); Tianjin Medical University Cancer Institute and Hospital, Tianjin, China (K Chen MD); Center for Molecular Epidemiology, National University of Singapore, Singapore (K S Chia MD); Qidong Cancer Registry, Qidong Liver Cancer Institute, Jiangsu, China (J G Chen MD); Hong Kong Cancer Registry, Hospital Authority, Hong Kong, China (S C K Law FRCR); Seoul Cancer Registry, Seoul National University College of Medicine, Seoul, South Korea (Y-O Ahn MD); Department of Epidemiology, Shanghai Cancer Institute, Shanghai, China (Y B Xiang MD); Bombay Cancer Registry, Indian Cancer Society, Mumbai, India (B B Yeole PhD); Busan Cancer Registry, National Cancer Center Research Institute, Kyonggi, South Korea (H R Shin MD); Incheon Cancer Registry, Inha University, Incheon, South Korea (Z H Woo MD); Lampang Cancer Registry, Lampang Cancer Center, Lampang, Thailand (N Martin MD); Department of Otolaryngology, Chiang Mai University, Chiang Mai, Thailand (Y Sumitsawan MD);

Songkhla Cancer Registry, Prince of Songkla University, Songkhla, Thailand (H Sriplung MD); Costa Rica National Tumor Registry, San Jose, Costa Rica (A O Barboza MD); Izmir Cancer Registry, Izmir Provincial Health Directorate, Alsancak, Turkey (S Eser MD); Barshi Cancer Registry, Nargis Dutt Memorial Cancer Hospital, Barshi, India (B M Nene MD); Cancer Unit, Srinagarind Hospital, Khon Kaen University, Khon Kaen, Thailand (K Suwanrungruang MSc); Natural Background Radiation Registry, Karunagappally, Kerala, India (P Jayalekshmi PhD); Tata Memorial Hospital, Mumbai, India (R Dikshit PhD); Kampala Cancer Registry, Makerere University Faculty of Medicine, Kampala, Uganda (H Wabinga MD); Rizal Cancer Registry, Rizal Medical Centre, Pasig City, Philippines (D B Esteban MD); Philippine Cancer Society, Manila Cancer Registry, Manila, Philippines (A Laudico MD); Karachi Cancer Registry, Liaquat National Medical College, Karachi, Pakistan (Y Bhurgri MD); National Cancer Registry of The Gambia, Banjul, The Gambia (E Bah MPH); and Faculty of Medicine, King Fahad Medical City, Riyadh, Saudi Arabia (N Al-Hamdan MD)

Correspondence to: Dr R Sankaranarayanan, Early Detection and Prevention Section, International Agency for Research on Cancer, Lyon, France sankar@iarc.fr

See Online for webappendix

in the target population.¹⁰ Uniform criteria³ and the disease codes of the International Classification of Diseases and Related Health Problems, tenth revision (ICD-10)¹¹ were used for coding the collected data.

An overview of this study is given in the figure. Participating registries initially submitted data during 2002–05, which was assessed for quality by IARC using standard methods and criteria.^{12,13} Validation checks were done for errors such as unusual combinations of cancer sites, morphology, sex, or age at diagnosis, and for inconsistencies and completeness of follow-up information. The registries were then contacted for clarifications and were advised on measures to improve data quality and on further active and passive follow-up methods to ascertain patients' vital status. Data were subsequently updated and received until 2008. 25 population-based cancer registries covering an entire nation, city, or rural district from 12 countries were included (table 1). 16 of the registries provided data on clinical extent of disease.

Patients

Of the 615 636 total cases of incident cancers (ICD-10: C00-96) registered in 1990–2001 from 25 cancer registries, a subset of 366 357 patients diagnosed with cancers of tongue, oral cavity, stomach, large bowel, larynx, lung, breast, cervix, ovary, and bladder were analysed in this study (figure). 13 registries contributed cases for ten cancer sites, and the others contributed data for one to nine sites (table 1). Cases excluded were those registered on the basis of information from a death certificate but with no traceable hospital record (11 733), those with no follow-up information on vital status after diagnosis (9584), and cases rejected on validation checks (3382). 341 658 cases were eligible for final analysis. The final diagnosis of these cases was mainly by histology, but some were diagnosed by clinical, biochemical, imaging, or endoscopic methods. The necessary minimum required data for this study are in the webappendix.

Procedures

The survival status of patients at or within 5 years from the date of diagnosis was obtained by active and passive methods (table 1).¹³ Passive follow-up relied on matching cancer cases with all-cause death information collected from death registration systems and hospital records using unique person numbers (in countries where such information is available—eg, South Korea) or by using a combination of personal identifiers from national population registers (first and last name, address and date of birth, etc) for record linkage. However, death registration is often incomplete in countries such as India, Philippines, Thailand, The Gambia, Uganda, and Pakistan, and assuming patients to be alive unless a death certificate was available would lead to a large underestimation of deaths. Therefore, in these countries, the following active measures were undertaken to establish the survival status of patients without death records: repeated visits to hospitals to

scrutinise clinical follow-up notes; visits to death-registry offices, churches, and mosques to collect information from their death registers; telephone or reply-paid postal enquiries; investigations in work places or neighbourhoods; and house visits for personal enquiries. Registries from countries that use passive methods, such as China, Costa Rica, Singapore, Saudi Arabia, and South Korea, were advised to use repeated matching with data sources to update follow-up.

The endpoint in this study was death, regardless of the cause. Survival for each cancer site (all clinical stages combined) is described in terms of 5-year age-standardised relative survival (ASRS). For countries providing data from more than one cancer registry, the median value of individual survival estimates and the minimum and maximum for the different registries are reported. Survival by clinical extent of disease at 5 years from diagnosis was reported for eight cancer sites; for any site, data from 10–15 registries in four to eight countries were used for analysis. Categories of clinical extent of disease were defined as follows: localised, for tumours confined to the organ of origin without invasion into the surrounding tissue or organ and without involvement of any regional or distant lymph nodes or organs; regional, for tumours that had invaded surrounding tissue or organ, with or without the involvement of the regional lymph nodes, but not involving non-regional lymph nodes or organs; distant metastasis, if the tumour had spread to the non-regional lymph nodes or distant organs; or unknown.

Statistical analysis

To limit chance variation, survival estimates were based on at least 100 cancer cases in total and at least 30 cases in each subgroup of country and cancer site. Given the paucity of data from Africa and the rarity of distant metastasis in patients with head and neck cancers, exceptions were made for the survival estimates reported from African countries and survival by clinical extent of disease for oral cavity and tongue cancers. Observed survival was computed by actuarial method¹⁴ following the semi-complete approach.¹⁵ Relative survival¹⁶ was calculated as the ratio of observed to expected survival in the general population of the same age and sex,¹⁷ to exclude the effect of competing causes of mortality and to facilitate survival comparisons between countries with different background mortalities. Expected survivals for individual registries were estimated from abridged life tables according to country, age, and sex.¹⁸ Relative survival was reported as ASRS to account for differences in the age of cancer cases.¹⁹ Estimated incident cancer cases from developing countries in 2002 were used as standard cancer populations.²⁰ Analyses were done using SAS software version 9.1.²¹

Role of the funding source

The funding source had no role in study design, data collection, analysis, interpretation, or writing of the article.

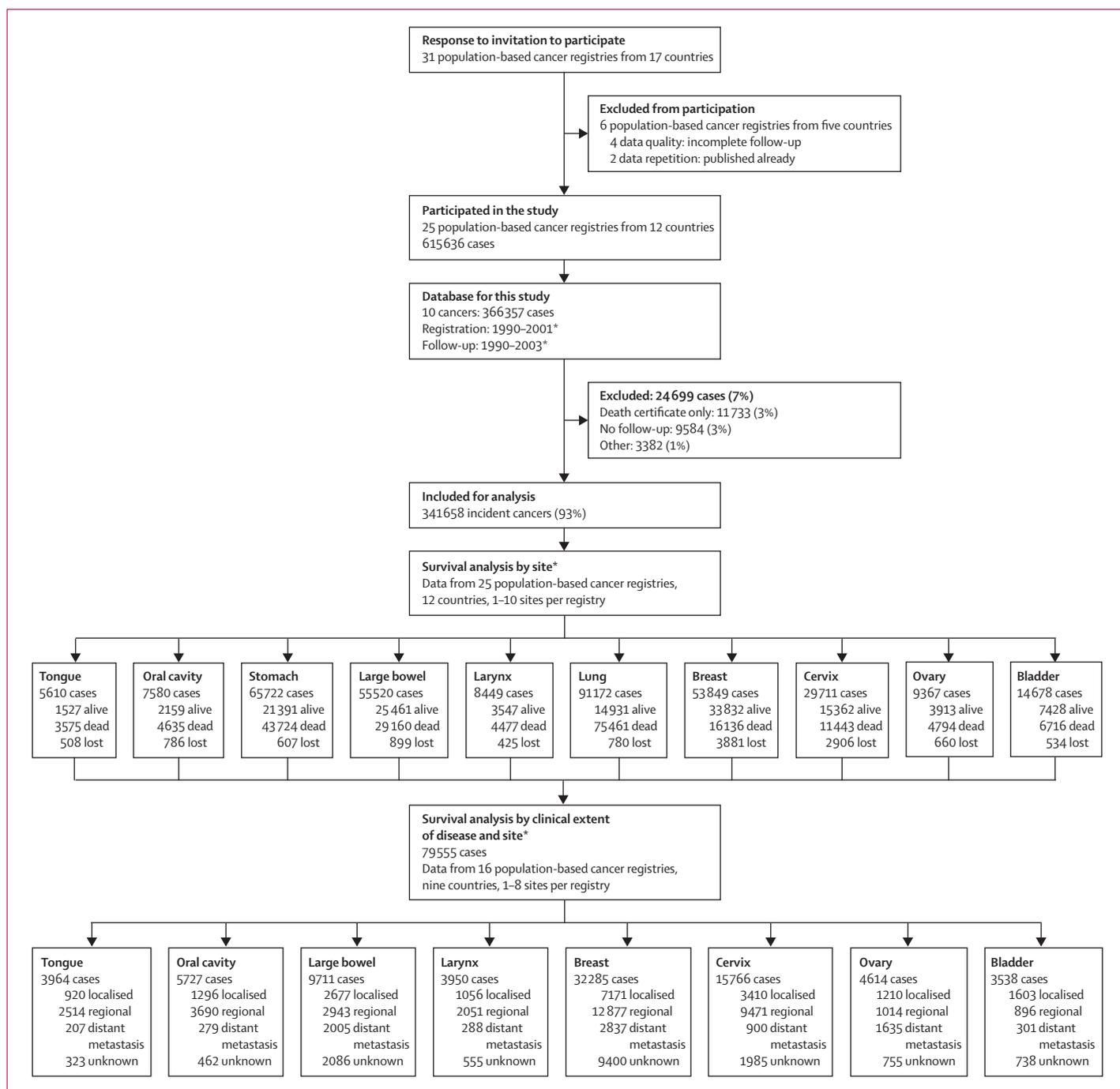


Figure: Overview of the survival study

*Varies for individual population-based cancer registries.

The corresponding author had full access to all study data and final responsibility for submission for publication.

Results

The range of values of data quality indices for different cancer sites by individual registries is given in table 1. Cases registered based on a death certificate only and

excluded from survival analysis ranged from 0 to 16%, and the percentage of registered cases included for survival analysis varied from 65 to 100%. Vital status (alive or dead) was known at the closing date of follow-up, or after 5 years from the date of diagnosis, for 30–100% of cases. Active efforts resulted in the availability or improvement of data on clinical extent of disease for

5–15% of cases and on vital status for 5–35% of cases in different registries. The number of patients included for survival analysis and median age, by cancer site and by country, are given in table 2. The figure shows the number of patients with vital status known on the last date of follow-up, the number lost to follow-up, and the number of patients in different clinical stages of disease.

Table 3 gives the 5-year ASRS (0–74 years) according to country for the ten cancer sites. In China, Singapore,

South Korea, and Turkey, the median relative survival rates were 76–82% for breast cancer, 63–79% for cervical cancer, 71–78% for bladder cancer, and 44–60% for large-bowel cancer. Survival was lowest in The Gambia, where ASRS did not exceed 22% for any cancer site, and in Uganda, where ASRS did not exceed 13% for any site except for breast cancer (46%). Comparing the maximum and minimum survival by registries in the same country showed marked variations for all cancers in China, most

	Registration period	Population	Follow-up		Number of sites	Data quality indices (% of cases)			
			Period	Method*		Death certificate only	Histological verification	Included for survival analysis	Complete follow-up at 5 years
Africa									
The Gambia	1993–97	National	1993–99	A	4	0–6	12–40	65–75	81–96
Uganda									
Kyadondo	1993–97	Mixed	1993–99	A	6	0–2	48–64	70–88	53–87
Asia									
China									
Hong Kong	1996–2001	Urban	1996–2003	P	10	0–1	77–94	92–96	100
Qidong	1992–2000	Rural	1992–2000	A+P	9	0–3	10–94	94–99	100
Shanghai	1992–95	Urban	1992–2000	P+A	10	0–1	40–93	99–100	100
Tianjin	1991–99	Urban	1991–2000	P+A	10	0–1	39–78	99–100	100
India									
Barshi	1993–2000	Rural	1993–2003	A	8	0–3	71–95	98–100	96–100
Bhopal	1991–95	Urban	1991–2000	A	10	0–2	60–90	66–89	100
Chennai	1990–99	Urban	1990–2001	A	9	0–5	64–86	86–94	83–92
Karunagappally	1991–97	Mixed	1991–99	A	10	0–8	50–94	89–100	97–100
Mumbai	1992–99	Urban	1992–2003	A	10	4–13	63–86	86–95	84–93
Pakistan									
Karachi	1995–99	Urban	1995–2003	A	2	0–1	98–99	92–93	72–74
Philippines									
Manila	1994–95	Urban	1994–2002	A+P	3	0–1	82–88	74–83	78–80
Rizal	1996–97	Urban	1996–2002	A+P	1	6	90	83	30
Saudi Arabia									
Riyadh	1994–96	Urban	1994–2001	P+A	1	0	99	93	80
Singapore									
	1993–97	National	1993–2001	P	10	0–2	82–100	91–98	100
South Korea									
Busan	1996–2001	Urban	1996–2003	P+A	10	2–12	64–93	85–97	100
Incheon	1997–2001	Urban	1997–2002	P+A	10	2–16	64–91	84–97	100
Seoul	1993–97	Urban	1993–2001	P+A	10	2–13	68–88	82–91	100
Thailand									
Chiang Mai	1993–97	Mixed	1993–2000	A	10	0–10	61–97	77–91	72–89
Khon Kaen	1993–97	Mixed	1993–2000	A+P	8	0–5	59–84	85–95	40–82
Lampang	1990–2000	Mixed	1990–2003	P+A	10	0–13	66–93	81–100	98–100
Songkhla	1990–99	Mixed	1990–2003	A	10	0–5	78–94	78–86	56–82
Turkey									
Izmir	1995–97	Mixed	1995–2003	A	6	0–1	88–96	98–99	82–87
Central America									
Costa Rica	1995–2000	National	1995–2003	P+A	2	2–3	91–92	82–86	74–83

P=passive follow-up. A=active follow-up. Mixed=urban+rural. *P+A=predominantly passive; A+P=predominantly active.

Table 1: Registration period, population covered, follow-up period and method, number of cancer sites, and data quality indices for registries in each country

	Tongue	Oral cavity	Stomach	Large bowel	Larynx	Lung	Breast	Cervix	Ovary	Bladder	Total
Africa											
The Gambia											
N	39	34	..	31	61	202*	367
MA	60	50	..	56	47	42
Uganda											
N	90	101	..	47	162	283	66	..	749
MA	54	53	..	48	46	43	40
Asia											
China											
N	1086	1128	28384	31762	2920	59917	21810	3883	2335	7610	160835
MA	60	65	68	68	67	68	51	61	55	70	..
E	10112	10112
India											
N	3376	4821	5098	3056	2763	5754	11013	10034	2967	1864	50746
MA	56	55	60	56	59	60	50	50	50	63	..
E	3307	4726	..	2943	2708	..	10812	9376	2932	1833	38637
Pakistan											
N	180	413	593
MA	55	50
E	180	413	593
Philippines											
N	248	1714	2339
MA	62	50
E	248	1714	2339
Saudi Arabia											
N	298	298
MA	45
E	298	298
Singapore											
N	120	135	2100	4254	311	4253	3204	984	718	496	16575
MA	61	65	68	66	67	69	50	52	48	69	..
E	120	135	..	4254	311	..	3204	984	718	496	10222
South Korea											
N	437	523	29156	13418	1488	15844	9442	8919	2317	3396	84940
MA	58	59	59	60	62	65	47	50	48	63	..
Thailand											
N	411	560	855	1932	489	5326	2354	3298	734	746	16705
MA	64	66	62	62	67	63	48	49	49	66	..
E	357	453	..	1551	453	..	2354	3298	734	643	9843
Turkey											
N	715	478	..	1329	246	230	566	3564
MA	62	59	..	52	52	55	64	..
E	715	478	..	1329	246	230	566	3564
Central America											
Costa Rica											
N	2462	1485	3947
MA	52	47
E	2462	1485	3947
Total											
N	5610	7580	65722	55520	8449	91172	53849	29711	9367	14678	341658
E	3964	5727	..	9711	3950	..	32285	15766	4614	3538	79555

N=total number of cases. MA=median age at diagnosis. E=number of cases included for extent-of-disease analysis. ..=data not available. *Random sample of total incident cases.

Table 2: Number of cases included for survival analysis, median age, and number included for extent-of-disease analysis by cancer site and country

	Tongue	Oral cavity	Stomach	Large bowel	Larynx	Lung	Breast	Cervix	Ovary	Bladder
Africa										
The Gambia	3	4	..	20	12	22
Uganda	0	8	..	0	46	13	9	..
Asia										
China	67 (64-68)	67 (44-71)	39 (20-44)	44 (36-63)	68 (49-73)	21 (6-32)	82 (58-90)	67 (48-79)	56 (35-67)	78 (47-81)
India	23 (12-30)	37 (26-45)	6 (4-15)	28 (6-31)	28 (16-38)	7 (1-13)	52 (31-54)	46 (34-60)	25 (19-29)	39 (10-48)
Pakistan	39	41
Philippines	40	47 (40-55)	37
Saudi Arabia	64
Singapore	44	49	27	52	66	9	76	66	62	72
South Korea	53 (52-60)	52 (48-54)	49 (46-50)	60 (57-64)	62 (59-76)	20 (14-21)	79 (78-81)	79 (76-79)	59 (54-62)	76 (72-79)
Thailand	32 (24-33)	36 (22-42)	12 (9-20)	35 (31-44)	36 (26-49)	9 (5-12)	63 (57-66)	61 (54-63)	47 (45-58)	48 (33-60)
Turkey	52	71	..	77	63	60	71
Central America										
Costa Rica	70	53

Data are median percentage of individual registries (minimum–maximum, if more than one registry). ..=data not available.

Table 3: 5-year age standardised (0–74 years) relative survival for cancers by country

cancers in India, a few cancers in Thailand, and laryngeal cancer in South Korea (table 3).

Observed 5-year survival and the proportion of cases presenting with different categories of clinical extent of disease were compared for countries with more developed versus less developed health-care services, for eight cancer sites (table 4). Data on clinical extent of disease were not available from any country in Africa. Among the countries with more developed health-care services, data on clinical extent of disease were available from Singapore for all sites, Turkey for most sites, and Hong Kong for breast cancer only. Available data on clinical extent of disease from these countries were pooled together for each site and classified as group A. Similarly, data on clinical extent of disease from countries with less developed health-care services were pooled together as group B. A decrease in survival with advanced stages of disease was seen for all cancers in both groups. For localised cancers of the breast, large bowel, larynx, ovary, and urinary bladder, and for all regional diseases at all sites, survival was higher among group A than group B.

Discussion

The countries in this study can be grouped into three tiers based on observed survival. Survival was highest in Hong Kong (data not included), South Korea, Singapore, and Turkey, where health services are well developed with advanced diagnostic and treatment centres and high per head GNI values. Survival was intermediate in Costa Rica, mainland China, Thailand, India, Saudi Arabia, Pakistan, and Philippines, where cancer health services are moderately developed with diagnostic and treatment facilities centred in and around urban cities, and with medium per head GNI.

Survival was lowest in The Gambia and Uganda, with poorly developed health services, as indicated by limited availability of cancer diagnostic and treatment facilities, and with very low per head GNI.

The level of development of health services and their efficiency to provide early diagnosis, treatment, and clinical follow-up care have a profound effect on cancer survival. We studied survival of patients with cancer in countries with a range of economic development levels. The interpretation of our results should take into account differences in the quality of data among registries in each country, and the wide differences in health awareness, socioeconomic development, human resources, health-services investment, and health-care accessibility between countries. Poorly developed and inaccessible health services result in inconsistencies in early diagnosis, adequate treatment, and follow-up care.

Survival results in our study were based on a sufficient number of cases, except for some cancer sites in patients in The Gambia and Uganda. Because the registries used passive or active methods of follow-up to varying extents, the possibility of overestimation of survival should be addressed. A recent study²² indicated that 5-year survival can be overestimated by up to 6% when ascertainment of deaths is up to 5% incomplete. Mortality ascertainment is likely to be incomplete with passive follow-up if the data linkage is not based on a unique personal identification number backed by a reliable death registration system. Therefore, an underestimation of deaths cannot be ruled out in some of the populations included in our study, despite our active efforts to ascertain such events.

A study of the implications of lack of active follow-up for various cancers in a registry in Chennai, India, showed an upward bias in population-based survival of 3–13% in

the presence of 5–21% random losses to follow-up.²³ This bias was estimated by assuming that patients lost to follow-up were alive at the closing date. For the registries that mainly used passive follow-up in our study, the overestimation of 5-year ASRS is likely to be no higher than 13%, if such misclassification of patients as alive at closing date does not exceed one in five.

In this study, a few registries using active methods reported substantial losses to follow up. However, studies using loss-adjusted survival methods with such data have shown that overestimation of population-based survival was small.^{24,25} We made considerable efforts to improve data quality for each registry; the survival estimates from our study are reliable and can serve as baseline for comparisons and for assessing improvements in survival outcome. The observed differences in survival between countries seem to be largely a result of differences in the availability and accessibility of early diagnosis and treatment, and, to a lesser extent, to data quality and reliability issues.

Populations followed up mainly with passive methods had consistently higher survival than those with mainly active follow-up. The registries in South Korea, Hong Kong, Singapore, and Costa Rica, that mostly used passive follow-up, had well developed health-information systems, with unique data linkage facilities and individual identity numbers or networking of hospitals. The wide availability and accessibility of screening programmes, early detection services, and cancer treatment facilities in these regions probably largely contributed to the high survival observed.

The survival patterns for regions in this study correlate well with the level of development and accessibility of health-care services. High survival rates for large-bowel, breast, cervical, and ovarian cancers in South Korea, Singapore, Izmir (Turkey), and Hong Kong are due to established screening and early detection programmes, and well developed and accessible diagnostic and treatment services in the public domain. The possibility of pathological misclassification between borderline and invasive ovarian cancers might partly explain the high survival observed for ovarian cancer in South Korea, China, Singapore, and Turkey. Cancer diagnosis and treatment services are underdeveloped in The Gambia and Uganda. Survival in most sub-Saharan African countries, and in some low-income countries in Asia and Central America that were not included in this study, is likely to be much lower than the survival observed in our study, due to the inadequacy or absence of cancer health services.

The large variations in survival within populations in regions of China, India, and Thailand reflect varying levels of development of cancer health services, particularly in urban versus rural areas. Higher survival in urban versus rural areas in India and China is the result of better cancer services in urban areas, with more accessibility for diagnosis and treatment. In China, survival for all cancers was consistently lower in the rural

	Pooled number of cases	5-year survival (frequency)			
		Localised	Regional	Distant metastasis	Unknown
Tongue					
Group A: Singapore	120	48.4 (26)	23.3 (25)	20.0 (4)	33.1 (45)
Group B: India, Pakistan, Thailand	3844	54.3 (23)	14.5 (65)	3.1 (5)	25.3 (7)
Oral cavity					
Group A: Singapore	135	52.3 (28)	26.5 (22)	.. (2)	28.9 (48)
Group B: India, Pakistan, Thailand	5592	60.2 (22)	23.8 (66)	3.3 (5)	28.8 (7)
Large bowel					
Group A: Singapore, Turkey	4969	64.1 (26)	45.7 (27)	8.6 (18)	41.8 (29)
Group B: India, Philippines, Thailand	4742	49.8 (29)	32.0 (34)	2.4 (23)	34.7 (14)
Larynx					
Group A: Singapore, Turkey	789	69.6 (34)	40.7 (19)	41.8 (5)	54.6 (42)
Group B: India, Thailand	3161	54.4 (25)	22.3 (60)	4.7 (8)	26.9 (7)
Breast					
Group A: China (Hong Kong), Singapore, Turkey	14 645	89.6 (17)	75.4 (32)	26.7 (2)	79.7 (49)
Group B: Costa Rica, India, Philippines, Saudi Arabia, Thailand	17 640	76.3 (26)	47.4 (47)	14.9 (14)	47.1 (13)
Cervix					
Group A: Singapore, Turkey	1230	69.5 (42)	52.2 (13)	18.6 (5)	57.5 (40)
Group B: Costa Rica, India, Philippines, Thailand	14 536	73.2 (20)	47.2 (64)	7.4 (6)	45.7 (10)
Ovary					
Group A: Singapore, Turkey	948	84.1 (40)	39.7 (4)	28.1 (27)	56.7 (29)
Group B: India, Thailand	3666	63.8 (22)	34.5 (27)	4.2 (38)	36.8 (13)
Bladder					
Group A: Singapore, Turkey	1062	61.3 (53)	34.8 (7)	16.4 (5)	54.0 (35)
Group B: India, Thailand	2476	43.8 (42)	24.9 (33)	2.3 (10)	35.6 (15)

Data are n or %. ..=data not available.

Table 4: 5-year survival and frequency distribution for cancers by clinical extent of disease, based on data from countries with more developed (group A) vs less developed (group B) health services

area of Qidong than in big cities like Shanghai and Tianjin (data not shown). In India, marked differences in survival for most cancers were noted between rural (Barshi), semi-urban (Karunagappally), or small urban (Bhopal) registries, whereas differences were small between the metropolitan cities of Chennai and Mumbai, where more developed and accessible health-care services are available (data not shown). In Thailand, the differences between registries were limited to a few cancers only. The data quality indices among registries within the same country did not show any major differences. Therefore, the differences in survival reflect differences in early diagnosis and the effect of inequalities in health-care infrastructure development, availability, and accessibility to health services. All three registries in South Korea showed no major differences in survival for any cancer (data not shown), possibly reflecting equitably developed and accessible health-care services across the country. The poor survival observed in The Gambia and Uganda emphasises the urgent need for government investments to improve health services.

China and South Korea, which reported the highest survival for most cancers, did not collect data on clinical extent of disease on a routine basis. Such information was available for selected cancer sites from Singapore and Turkey, with high survival reported, and for Costa Rica, India, Pakistan, Philippines, Saudi Arabia, and Thailand, with intermediate survival. Since data collection was from heterogeneous sources, misclassification of extent of disease, especially in localised or regional categories, was possible. Also, using data from specialised and non-specialised medical institutions could lead to unrealistic comparisons between registries. To minimise this problem, we classified countries into two groups based on survival results and the level of development of cancer health care (according to per head GNI values). The pooled data for each group showed decreasing survival with increasing stage at presentation for all cancers, which added strength to the data quality in this study. The higher survival observed for group A countries for localised large-bowel, larynx, breast, ovarian, and bladder cancers (which largely require radical surgical treatment) and for regional disease (which require multimodal treatment) are largely due to the differences in the development and accessibility of diagnostic and treatment services. The high survival observed for advanced (distant metastasis) laryngeal cancer in Singapore and Turkey might reflect misclassification between regional and distant metastatic disease.

This study would not have been possible without the availability of reliable population-based cancer registries. It is important to organise such information systems in regions that lack them. It is likely that survival in many low-income and middle-income countries that were not included in this study, particularly in sub-Saharan Africa, would be lower than reported here. The striking differences in cancer survival between countries emphasises the need for urgent and adequate investments in comprehensive cancer control. In addition to equitable accessibility to health services, efforts are needed to improve public and professional awareness, early detection and prompt treatment using locally feasible yet effective regimens, health-services infrastructure, human-resources development, and referral pathways. In many sub-Saharan African countries, such investments in health services have never been made, and national cancer-control policies should focus on balanced investments in prevention and early detection and treatment of common cancers. Further study of survival patterns in developing countries should address the limitations for follow-up and devise strategies to improve completeness by use of complementary active methods. In conclusion, our results largely reflect the differences in availability and accessibility of health care. These data can be used as the baseline for sustainable improvements in the

future, through adequate and sincere investments in improving awareness, health-services infrastructure, and accessibility.

Contributors

RSa was responsible for the conception and design of the study, and the monitoring, supervision, acquisition, analysis, and interpretation of data. RSw was responsible for the conduct of the study and the acquisition, analysis, and interpretation of data. HB was responsible for the conduct of the study and the analysis and interpretation of data. KC, KSC, JGC, SCKL, Y-OA, YBX, BBY, HRS, VS, ZHW, NM, YS, HS, AOB, SE, BMN, KS, PJ, RD, HW, DBE, AL, YB, EB, and NA-H were responsible for the conduct of the study and acquisition of the data in their respective cancer registries. All authors were involved in drafting of the manuscript.

Conflicts of interest

The authors declared no conflicts of interest.

Acknowledgments

We are grateful to the following organisations for funding support to cancer survival studies: Association for International Cancer Research (AICR; St Andrews, UK), Association pour la Recherche sur le Cancer (ARC, Villejuif, France), and the Bill & Melinda Gates Foundation (Seattle, USA). Final analysis was carried out during the post-doctoral fellowship awarded by IARC to RSw. We thank Evelyn Bayle, IARC, for her help in the preparation of the manuscript.

References

- 1 Parkin DM, Whelan SL, Ferlay J, Storm H, eds. Cancer incidence in five continents, Vol I–III. IARC Cancerbase No 7. Lyon, France: International Agency for Research on Cancer, 2005.
- 2 Curado MP, Edwards B, Shin HR, et al, eds. Cancer incidence in five continents, Vol IX. IARC Scientific Publications No 160. Lyon, France: International Agency for Research on Cancer, 2007.
- 3 Sant M, Allemani C, Santaquilani M, et al. EUROCARE-4 survival of cancer patients diagnosed in 1995–1999: results and commentary. *Eur J Cancer* 2009; 45: 931–91.
- 4 Ries LAG, Harkins D, Krapcho M, et al, eds. SEER Cancer Statistics Review, 1975–2003. Bethesda, MD, USA: National Cancer Institute, 2006.
- 5 Coleman MP, Quaresma M, Berrino F, et al. Cancer survival in five continents: a worldwide population-based study (CONCORD). *Lancet Oncol* 2008; 9: 730–56.
- 6 Tsukuma H, Ajiki W, Ioka A, Oshima A, research group of population-based cancer registries of Japan. Survival of cancer patients diagnosed between 1993 and 1996: a collaborative study of population-based cancer registries in Japan. *Jpn J Clin Oncol* 2006; 36: 602–07.
- 7 Kune GA, Kune S, Field B, et al. Survival in patients with large-bowel cancer. a population-based investigation from the Melbourne Colorectal Cancer Study. *Dis Colon Rectum* 1990; 33: 938–46.
- 8 Sankaranarayanan R, Black RJ, Parkin DM, eds. Cancer survival in developing countries. IARC Scientific Publications No 145. Lyon, France: International Agency for Research on Cancer, 1998.
- 9 Sankaranarayanan R, Swaminathan R, Black RJ. Global variations in cancer survival: study group on cancer survival in developing countries. *Cancer* 1996; 78: 2461–64.
- 10 Jensen OM, Parkin DM, MacLennan R, Muir CS, Skeet RG, eds. Cancer registration principles and methods. IARC Scientific Publications No 95. Lyon, France: International Agency for Research on Cancer, 1999.
- 11 WHO. International statistical classification of diseases and related health problems, tenth revision (ICD-10), Vol 1. Geneva, Switzerland: World Health Organization, 1992.
- 12 Berrino F, Sant M, Verdecchia A, eds. Survival of cancer patients in Europe: the EUROCARE Study. IARC Scientific Publications No 132. Lyon, France: International Agency for Research on Cancer, 1995.
- 13 Swaminathan R, Black RJ, Sankaranarayanan R. Database on cancer survival from developing countries. In: Sankaranarayanan R, Black RJ, Parkin DM, eds. IARC Scientific Publications No 145. Lyon, France: International Agency for Research on Cancer, 1998.

- 14 Cutler SJ, Ederer F. Maximum utilization of the life table method in analysing survival. *J Chron Dis* 1958; **8**: 699–712.
- 15 Brenner H, Gefeller O. An alternative approach to monitoring cancer patient survival. *Cancer* 1996; **78**: 2004–10.
- 16 Ederer F, Axtell LM, Cutler SJ. The relative survival rate: a statistical methodology. *Monogr Natl Cancer Inst* 1961; **6**: 101–21.
- 17 Hakulinen T. Cancer survival corrected for heterogeneity in patient withdrawal. *Biometrics* 1982; **38**: 933–42.
- 18 Lopez AD, Ahmad OB, Guillot M, Inoue M, Ferguson BD, Salomon JA. Life tables for 191 countries for 2000: Data, Methods, Results (GPE discussion paper No. 40). Health Systems Performance Assessments Peer Review, Technical Documentation IV Outcomes: Population Health. Evidence and Information for Policy (EIP). Geneva, Switzerland: World Health Organization, 2001.
- 19 Brenner H, Arndt V, Gefeller O, Hakulinen T. An alternative approach to age adjustment of cancer survival rates. *Eur J Cancer* 2004; **40**: 2317–22.
- 20 Ferlay J, Bray F, Pisani P, Parkin DM. GLOBOCAN 2002: cancer incidence, mortality and prevalence worldwide. IARC CancerBase No 5, version 2.0. Lyon, France: IARC Press, 2004.
- 21 Brenner H, Gefeller O, Hakulinen T. A computer program for period analysis of cancer patient survival. *Eur J Cancer* 2002; **38**: 690–95.
- 22 Brenner H, Hakulinen T. Implications of incomplete registration of deaths on long-term survival estimates from population-based cancer registries. *Int J Cancer* 2009; **125**: 432–37.
- 23 Swaminathan R, Rama R, Shanta V. Lack of active follow up of cancer patients in Chennai, India: implications for population based survival estimates. *Bull World Health Organ* 2008; **86**: 509–15.
- 24 Sriamporn S, Swaminathan R, Parkin DM, Kamsa-Ard S, Hakama M. Loss-adjusted survival of cervix cancer in Khon Kaen, Northeast Thailand. *B J Cancer* 2004; **91**: 106–10.
- 25 Swaminathan R, Sankaranarayanan R, Hakama M, Shanta V. Effect of loss to follow up on population based cancer survival rates in developing countries. *Int J Cancer* 2002; **13**: 172.

Lack of active follow-up of cancer patients in Chennai, India: implications for population-based survival estimates

Rajaraman Swaminathan,^a Ranganathan Rama,^a & Viswanathan Shanta^a

Objective To measure the bias in absolute cancer survival estimates in the absence of active follow-up of cancer patients in developing countries.

Methods Included in the study were all incident cases of the 10 most common cancers and corresponding subtypes plus all tobacco-related cancers not ranked among the top 10 that were registered in the population-based cancer registry in Chennai, India, during 1990–1999 and followed through 2001. Registered incident cases were first matched with those in the all-cause mortality database from the vital statistics division of the Corporation of Chennai. Unmatched incident cancer cases were then actively followed up to determine their survival status. Absolute survival was estimated by using an actuarial method and applying different assumptions regarding the survival status (alive/dead) of cases under passive and active follow-up.

Findings Before active follow-up, matches between cases ranged from 20% to 66%, depending on the site of the primary tumour. Active follow-up of unmatched incident cases revealed that 15% to 43% had died by the end of the follow-up period, while the survival status of 4% to 38% remained unknown. Before active follow-up of cancer patients, 5-year absolute survival was estimated to be between 22% and 47% higher, than when conventional actuarial assumption methods were applied to cases that were lost to follow-up. The smallest survival estimates were obtained when cases lost to follow-up were excluded from the analysis.

Conclusion Under the conditions that prevail in India and other developing countries, active follow-up of cancer patients yields the most reliable estimates of cancer survival rates. Passive case follow-up alone or applying standard methods to estimate survival is likely to result in an upward bias.

Bulletin of the World Health Organization 2008;86:509–515.

Une traduction en français de ce résumé figure à la fin de l'article. Al final del artículo se facilita una traducción al español. الترجمة العربية لهذه الخلاصة في نهاية النص الكامل لهذه المقالة.

Introduction

In recent decades, incident cancer cases have been systematically and continuously registered all over the world using both active and passive methods. Passive registration methods, which may or may not be facilitated by the law, are those in which incident cancer cases are notified and the data are involuntarily received by the registry from the respective sources. Active cancer registration methods consist of collecting data from other sources voluntarily. Data from 53 registries in 25 developing countries were published in 2002 by the International Agency for Research on Cancer in Lyon, France.¹ Cancer was a notifiable disease in 49% of the 53 registries, while data on incident cancers were collected entirely by passive methods in 34%. In less than

one-third of the registries practising passive registration, data linkages were based on unique identification numbers.¹

In India, cancer is not a notifiable disease. Hence, cancer cases are primarily registered through active methods.^{2–6} The population-based cancer registry (PBCR) in Chennai, known as the Madras Metropolitan Tumour Registry (MMTR), is based at the Cancer Institute (Women's India Association) and has been a part of the National Cancer Registry Program of the Indian Council of Medical Research, a government entity, since 1981.

Official cancer mortality data from the vital statistics division is generally integrated into the PBCR. However, in most developing countries, including India, death certificates are often inaccurate, so that all-cause mortality data

should be used to supplement cancer mortality statistics.⁷

Having reliable information on survival from cancer has long been recognized as important for cancer control activities. Monitoring population-based survival rates is useful for patient care and health care planning. Such rates are free from case selection bias and reflect average cancer-related outcomes in a given region. Population-based cancer survival estimates have been increasingly available in developing countries since the early 1990s, but at least one-third of them are based exclusively on passive follow-up.⁸ The present study aims to measure the bias resulting from absolute survival estimates in the absence of active case follow-up and when different assumptions are made regarding the survival status of cancer patients in developing countries.

^a Division of Epidemiology and Cancer Registry, Cancer Institute (Women's India Association), 38 Sardar Patel Road, Chennai 600036, India.

Correspondence to Rajaraman Swaminathan (e-mail: iarcsurvival@yahoo.co.uk).

doi:10.2471/BLT.07.046979

(Submitted: 18 August 2007 – Final revised version received: 23 November 2007 – Accepted: 26 November 2007 – Published online: 20 December 2007)

Table 1. Survival status of incident cancer cases registered in 1990–1999 and followed through 2001, PBCR, Chennai, India

Tumour site/type	Cases included in survival analysis	Passive follow-up		Active follow-up	
		Matched deaths (%)	Additional deaths identified (%)	Cases alive at closing date (%)	Survival status unknown at closing date (%)
Lip	86	19.8	33.7	11.6	34.9
Tongue	988	37.6	32.5	5.5	24.4
Oral cavity	1662	31.8	31.5	10.2	26.5
Tonsil	250	42.8	42.8	6.4	8.0
Hypopharynx	1017	41.4	40.5	5.8	12.3
Oesophagus	2016	51.0	36.3	2.9	9.8
Stomach	2681	51.9	33.0	4.5	10.6
Pancreas	328	57.9	30.8	7.0	4.3
Larynx	722	40.2	23.0	19.6	17.2
Lung	1806	59.2	28.0	2.4	10.4
Breast	3067	28.5	20.0	28.2	23.3
Cervix	4438	25.5	16.7	19.8	38.0
Ovary	808	39.7	20.5	17.2	22.6
Urinary bladder	442	38.9	30.1	14.0	17.0
Hodgkin lymphoma	298	30.9	26.5	24.8	17.8
Non-Hodgkin lymphoma	868	44.1	25.2	15.0	15.7
Lymphoid leukaemia	433	45.5	29.1	11.3	14.1
Myeloid leukaemia	465	59.6	18.9	7.5	14.0
Leukaemia, type unspecified	85	65.9	15.3	5.9	12.9

PBCR, population-based cancer registry.

Methods

Included in the study were all incident cases of the 10 most common broadly-defined cancers and corresponding subtypes (for cancers of the oral cavity, lymphomas and leukaemias), plus tobacco-related cancers not ranked among the top 10 (such as pancreas and urinary bladder), that were registered in the MMTR in Chennai during 1990–1999 and followed through 31 December 2001.

Data on incident cancer cases in the MMTR were obtained by direct interview of patients by cancer registrars at selected source hospitals at the time of registration and/or by perusal of medical records at those hospitals using a validated, standardized questionnaire common to all registries in India. Interviewers were trained by senior investigators of the registry project at the base institution where the registry is physically located.³ Data on cancer deaths through 1991 and on all-cause mortality since 1992 were extracted from death certificates maintained at the vital statistics division of the Corporation of Chennai.^{3,7} Incident cancer cases in the MMTR were then matched with cases in the mortality da-

tabase primarily using each individual's personal identity details. Cancer cases for which no matches were found in the mortality database were actively followed to determine their survival status. Medical records at source hospitals that imposed restrictions on active follow-up were examined once every 3 years or less in order to track patients' attendance at clinical follow-up visits. Postal or telephone enquiries among patients or their relatives and friends and other contacts were carried out by cured cancer patients from the locality, volunteer service organizations, and health workers. House visits, which make it possible to interrogate neighbourhood residents, are the most common active follow-up method pursued by patient registries in India to effectively determine the survival status of patients who have migrated (common in urban areas).

Different actuarial assumptions on the survival status of subjects were made during follow-up for the purpose of this study. Subjects were designated as belonging to the following categories: (A) when they were matched with mortality data obtained by routine registry data linkage with official mortality statistics without any active follow-up;

(B) when they could not be matched through routine registry data linkage with official mortality statistics and their death was ascertained through active follow-up; (C) when they were lost to follow-up but known to be alive until a specific date, with unknown survival status at the close of follow-up; and (D) when they had completed follow-up and were known to be alive on the closing date.

The follow-up status was classified into four different case scenarios depending on the assumptions made, as follows:

Case 1: Passive follow-up only of cancer cases not matched with official mortality data but presumed to be alive at the close of follow-up. In this scenario, subjects in category A were treated as having died on their respective dates of death, while subjects B, C, and D were treated as having been alive on the last day of follow-up.

Case 2: Passive and active follow-up, with cases lost to follow-up presumed to be alive on the last day of follow-up. In this scenario, subjects A and B were treated as having died on their respective dates of demise, while subjects C and D were treated as having been alive on the last day of follow-up.

Table 2. Incident cancer cases included in the survival analysis, among those registered in 1990–1999 and followed through 2001, PBCR, Chennai, India

Tumour site/type	Number of cases included in survival analysis									
	Total	Passive follow-up only		Passive and active follow-up						
		Case 1 ^a		Case 2 ^b		Case 3 ^c			Case 4 ^d	
		Dead	Presumed alive at closing date	Dead	Presumed alive at closing date	Dead	Alive	Lost to follow-up	Dead	Alive
Lip	86	17	69	46	40	46	10	30	46	10
Tongue	988	371	617	693	295	693	54	241	693	54
Oral cavity	1662	528	1134	1052	610	1052	169	441	1052	169
Tonsil	250	107	143	214	36	214	16	20	214	16
Hypopharynx	1017	421	596	833	184	833	59	125	833	59
Oesophagus	2016	1028	988	1759	257	1759	59	198	1759	59
Stomach	2681	1392	1289	2277	404	2277	120	284	2277	120
Pancreas	328	190	138	291	37	291	23	14	291	23
Larynx	722	290	432	456	266	456	142	124	456	142
Lung	1806	1069	737	1574	232	1574	45	187	1574	45
Breast	3067	875	2192	1489	1578	1489	862	716	1489	862
Cervix	4438	1131	3307	1874	2564	1874	878	1686	1874	878
Ovary	808	321	487	487	321	487	138	183	487	138
Urinary bladder	442	172	270	305	137	305	62	75	305	62
Hodgkin lymphoma	298	92	206	171	127	171	74	53	171	74
Non-Hodgkin lymphoma	868	383	485	602	266	602	130	136	602	130
Lymphoid leukaemia	433	197	236	323	110	323	49	61	323	49
Myeloid leukaemia	465	277	188	365	100	365	35	65	365	35
Leukaemia, type unspecified	85	56	29	69	16	69	5	11	69	5

PBCR, population-based cancer registry.

^a Case 1: Passive follow-up only, with cancer cases not matched with those in the official mortality database presumed to be alive on the closing date.

^b Case 2: Passive and active follow-up, with cases lost to follow-up presumed to be alive on the closing date.

^c Case 3: Passive and active follow-up, with cases lost to follow-up censored on the last date their survival status was known.

^d Case 4: Passive and active follow-up, with cases lost to follow-up excluded from survival analysis.

Case 3: Passive and active follow-up, with cases lost to follow-up censored on the last date on which their survival status was known. Under this case scenario, subjects A and B were treated as having died on their respective dates of demise; subjects in category D were treated as having been alive on the last day of follow-up, and subjects in category C were treated as having been alive until a specific date and censored thereafter for the survival analysis, based on actuarial assumption.

Case 4: Passive and active follow-up, with cases lost to follow-up excluded from the survival analysis. This resembles Case 3, excepting that subjects in category C were excluded from the survival analysis.

Absolute survival probability, also known as crude survival, was estimated through an actuarial approach.⁹

However, the assumptions made in this study differed from those normally made using the routine actuarial method.

Findings

Table 1 gives the survival status of incident cancer cases, for primary tumours of different types, in accordance with the follow-up method used. Deaths in the all-cause mortality database that were matched with cases in the incident cancer database without any active follow-up ranged between 20% (lip cancer) and 66% (leukaemias, type unspecified). Of those cancer cases having no match in the mortality database and actively followed, 15% (leukaemia, type unspecified) to 43% (cancer of the tonsil) had died, and 3% (oesophageal cancer) to 28% (female breast cancer) were alive by the end of the follow-up

period. Survival status was unknown in 4% (pancreatic cancer) to 38% (cervical cancer) of the cases on the last day of follow-up. As shown in Table 2, a variable number of cases, depending on survival status, was used to estimate absolute survival under different actuarial assumptions at follow-up.

Table 3 shows the frequency (%) of losses to follow-up at varying time intervals from the time of diagnosis: < 1 year, 1–3 years, 3–5 years and > 5 years. This information can be obtained only through active follow-up. For most primary tumour sites, the highest proportion of losses to follow-up occurred within the first year from diagnosis, with figures ranging from 3% for lymphoid leukaemia to 15% for ovarian cancer cases. From about 1% of pancreatic cancer to 26% of lip cancer cases were lost to follow-up after 5 years from diagnosis. Very small proportions

were lost to follow-up between 1–3 years and 3–5 years from diagnosis.

Table 4 gives the 5-year absolute survival (%) estimated by actuarial methods under different assumptions on the survival status of subjects that were followed passively, actively, or both. The differences in 5-year absolute survival, in percentages, between cases 1 and 2 were smallest among cases of leukaemia (type unspecified) (15.1%), cervical cancer (16.5%), and myeloid leukaemia (18.9%), and highest among patients with cancers of the tonsil (41.3%), hypopharynx (39.2%), and lip (34.9%). In the absence of active follow-up (case 1), 5-year absolute survival was estimated to be higher by 22% (leukaemia, type unspecified) to 47% (hypopharyngeal cancer) than when cases were actively followed and were lost to follow-up at a known point in time (case 3). In relative terms, odds ratios (OR) reflecting survival differences were largest for oesophageal cancer (OR: 12.9) and smallest for leukaemia (type unspecified) (OR: 4.0). Cases 2 and 4 represent the two extremes of a survival spectrum, with the actuarial estimate assuming random withdrawal falling somewhere in between. The more losses to follow-up, the greater the uncertainty and potential for bias in the actuarial estimate. The absolute differences in 5-year survival between cases 2 and 4 were substantial for cancers of the tongue (13.8%) and ovary (18.4%).

Discussion

Survival estimates of unselected groups of cancer patients from population-based cancer registries can serve as an important index for evaluating cancer diagnosis and treatment and the effectiveness of overall cancer services in a given region.⁸ Of the 53 registries from 25 developing countries that published data on cancer incidence and mortality in 2002, less than half have published data on cancer survival despite their long history of cancer registration.^{1,8} In India, only six out of more than 20 registries have undertaken survival studies.^{2,8}

Unlike mortality data collection, follow-up is not usually integrated with routine population-based cancer registration practices. In most developed countries, passive follow-up of cancer patients is carried out through

Table 3. Distribution of incident cancer cases lost to follow-up, among those registered in 1990–1999 and followed through 2001, PBCR, Chennai, India

Tumour site/type	Losses to follow-up by years from diagnosis (%)			
	< 1	1–3	3–5	> 5
Lip	7.0	2.3	0.0	25.6
Tongue	13.1	2.6	1.2	7.5
Oral cavity	10.3	2.2	1.8	12.2
Tonsil	4.8	0.8	0.0	12.4
Hypopharynx	9.0	0.6	0.0	2.3
Oesophagus	6.7	0.9	0.4	1.8
Stomach	7.3	0.9	0.9	1.5
Pancreas	3.1	0.3	0.3	0.6
Larynx	6.7	0.8	0.3	9.4
Lung	8.0	0.7	0.3	1.4
Breast	12.4	2.9	2.0	6.0
Cervix	11.0	3.7	2.5	20.8
Ovary	14.7	4.6	1.4	1.9
Urinary bladder	10.9	1.6	0.2	4.3
Hodgkin lymphoma	6.4	1.7	1.0	8.7
Non-Hodgkin lymphoma	10.9	1.3	0.6	2.9
Lymphoid leukaemia	2.8	3.2	3.5	4.6
Myeloid leukaemia	8.6	1.3	0.4	3.7
Leukaemia, type unspecified	10.5	0.0	0.0	2.4

PBCR, population-based cancer registry.

the use of a personal identification number (PIN) matched with mortality databases. In making survival analyses, cancer cases are presumed to be alive when no information on death has been traced by a particular reference date. For losses to follow-up, non-informative or random censoring is anticipated (i.e. the losses to follow-up are assumed to be independent of the risk of death). However, in most developing countries, including India, unique citizen identifiers (such as PINs) do not exist; mortality registration systems, especially medical certification of deaths, are deficient, and the identity particulars of deceased individuals are often inaccurate. Thus, passive means of follow-up alone may not be sufficient to perform a meaningful survival analysis.

Ten registries from five developing countries contributed data on survival for the first time to the International Agency for Research on Cancer monograph on *Cancer survival in developing countries*,⁸ and four of them (Qidong and Shanghai registries from China; Cuba; and Rizal from the Philippines) relied either entirely or predominantly on passive follow-up methods. All four registries from India (Bangalore, Barshi, Bombay and Madras) that contributed data to that monograph had employed

active follow-up. In the forthcoming second volume of the same publication, many more registries submitted data on survival and several of them adhered to passive methods of follow-up. Thus, active methods are needed and the effect of passive registry follow-up on survival estimates should be ascertained. The authors have done this by using data from the Chennai registry in India and generalizing their conclusions to other developing countries.

The Chennai registry has collected data on all-cause mortality from the vital statistics division of the Corporation of Chennai since 1992. The general mortality-to-cancer incidence ratio was 45% in 1992–2001 and 23% before 1992, when only cancer mortality data were available.⁷ However, this did not account for all the deaths that had occurred among the incident cancer cases in the Chennai cancer registry. The active follow-up of cancer cases that could not be matched with cases in the all-cause mortality database revealed additional deaths, ranging from 15% more deaths among patients with leukaemia (type unspecified) to 43% more deaths among patients with cancer of the tonsil. The main reasons deaths could not be unambiguously matched with cases in the cancer registry data-

base were: (i) incomplete identity information about the deceased in death certificates/records; (ii) migration of cases within the registry area before death, and (iii) inaccurate details given by persons reporting the death. These factors are difficult to overcome despite the full availability of cause-specific mortality data in the region under study.

If invalid actuarial assumptions are made, deaths are underreported and the impact on absolute survival is large. Studies from developed countries employing unique case identifiers to link data passively have acknowledged the need to correct for survival status (alive/dead) through active follow-up, as well as the potential impact of active follow-up on survival.^{10,11}

In our study, losses to follow-up were most frequent within 1 year of diagnosis.^{12–16} A different pattern has been observed in Thailand, with the highest losses occurring more than 5 years from diagnosis.⁸ Losses to follow-up at varying times thus affect actuarial survival estimates under passive follow-up. The highest dropout rates within the first year of cancer diagnosis are often due to death, while the long-term losses to follow-up occur mainly among survivors. Many studies exclude cases that are lost to follow-up from survival analyses.^{8,13,15} As shown by our case 4 scenario, such exclusions may result in a substantial bias whose magnitude depends on the number of losses to follow-up, with losses not occurring randomly or independently of the risk of death. Loss-adjusted survival methods have been proposed¹⁷ and applied to survival studies, with many losses to follow-up considered non-random.^{13,18} After adjusting for cases lost to follow-up in these studies, only minimal differences were noted, ranging from 1% to 5% based on the data obtained from the population-based cancer registry, indicating that the losses were practically random. However, the same could not be said of survival studies using hospital cancer registry data, with differences in the order of 15%.^{13,17} These differences typically represent the advantages of using population-based cancer registry data rather than hospital series.

The study clearly shows that in a population-based cancer registry series, passive follow-up, as represented by our case 1 approach, is unidirectional

Table 4. Five-year absolute survival under different assumptions regarding survival status among incident cancer cases registered in 1990–1999 and followed through 2001, PBCR, Chennai, India

Tumour site/type	5-year absolute survival (%)			
	Passive follow-up	Active follow-up		
		Case 1 ^a	Case 2 ^b	Case 3 ^c
Lip	79.5	44.6	40.7	39.5
Tongue	62.1	29.2	19.4	15.4
Oral cavity	68.5	37.1	30.5	26.4
Tonsil	58.5	17.2	13.7	10.8
Hypopharynx	59.2	20.0	12.5	9.6
Oesophagus	48.9	12.9	6.9	5.0
Stomach	47.9	15.0	8.6	5.6
Pancreas	41.8	10.9	7.9	6.5
Larynx	59.0	35.1	30.7	28.4
Lung	40.8	13.2	6.5	4.2
Breast	71.6	51.5	43.7	39.6
Cervix	75.5	59.0	54.0	49.4
Ovary	60.1	39.5	27.4	21.1
Urinary bladder	61.3	31.0	23.2	20.0
Hodgkin lymphoma	69.1	42.6	39.4	35.9
Non-Hodgkin lymphoma	55.6	29.7	21.6	16.8
Lymphoid leukaemia	54.3	26.5	23.8	15.5
Myeloid leukaemia	40.4	21.5	14.7	10.9
Leukaemia, type unspecified	32.9	17.8	10.9	6.2

PBCR, population-based cancer registry.

^a Case 1: Passive follow-up only, with cancer cases not matched with those in the official mortality database presumed to be alive on the closing date.

^b Case 2: Passive and active follow-up, with cases lost to follow-up presumed to be alive on the closing date.

^c Case 3: Passive and active follow-up, with cases lost to follow-up censored on the last date their survival status was known.

^d Case 4: Passive and active follow-up, with cases lost to follow-up excluded from survival analysis.

and leads to potentially biased survival estimates. Our case 3 scenario – applying an actuarial approach after improving the follow-up data by using an active method – provides a closer estimate of true survival. Cases 2 and 4 yield the largest and smallest residual bias, respectively, when the follow-up data ascertained by the active method is incomplete. Using a loss-adjusted survival approach is meaningless if the missing data is associated with the risk of death and with prognostic factors. A more complete analysis would bring out whether true differences existed between the four case scenarios.

Conclusion

Under the conditions that prevail in India and other developing countries, with incomplete mortality registration, no unique case identifiers for linking data and poor health information systems, active follow-up of cancer patients

yields the most reliable estimates of cancer survival rates. Passive follow-up alone and standard methods of estimating survival are likely to result in an upward bias. ■

Acknowledgements

The authors thank the registry medical officer, statistician, social investigators, computer programmer and data entry operators who were responsible for data collection, entry and processing. They also thank the source hospitals and vital statistics division that provided data to the registry.

Funding: The population-based cancer registry is partly funded by the Indian Council of Medical Research and the survival project was fully funded by the Screening Group of the International Agency for Research on Cancer in Lyon, France.

Competing interests: None declared.

Résumé

Manque de suivi actif des patients cancéreux à Chennai, en Inde: implications pour les estimations du taux de survie en population

Objectif Mesurer le biais affectant les estimations du taux de survie absolue au cancer en l'absence de suivi actif des patients cancéreux dans les pays en développement.

Méthodes Ont été inclus dans l'étude tous les cas incidents des 10 cancers les plus courants et des sous-types correspondants, plus tous les cancers liés au tabac non classés parmi les 10 premiers recensés dans le registre des cancers de la population de Chennai, en Inde, au cours de la période 1990-1999, et suivis jusqu'en 2001. Les cas incidents enregistrés ont d'abord été appariés avec ceux figurant dans la base de données de mortalité toutes causes confondues de la division statistiques vitales de la Corporation de Chennai. Les cas de cancer incidents non appariés ont ensuite fait l'objet d'un suivi actif pour déterminer leur statut de survie. Le taux de survie absolue a été estimé en utilisant une méthode actuarielle et en appliquant différentes hypothèses concernant le statut de survie (vivant/mort) des cas, dans les situations de suivi passif et actif.

Résultats Avant le suivi actif, l'appariement obtenu allait de 20 à 66 %, selon le site de la tumeur primaire. Un suivi actif des cas incidents non appariés a révélé que 15 à 43 % d'entre eux étaient décédés à la fin de la période de suivi et que le statut de survie de 4 à 38 % de ces cas restait inconnu. Avant le suivi actif des patients cancéreux, on estimait que le taux de survie absolue à 5 ans se situait entre 22 et 47 %, soit plus qu'après l'application aux cas perdus pour le suivi de méthodes actuarielles hypothétiques classiques. Les estimations les plus faibles des taux de survie ont été obtenues en excluant les cas perdus pour le suivi de l'analyse.

Conclusion Dans les conditions qui prévalent en Inde et dans d'autres pays en développement, le suivi actif des patients cancéreux fournit les estimations les plus fiables des taux de survie au cancer. Le suivi passif seul ou l'application de méthodes classiques pour estimer la survie sont susceptibles d'entraîner un biais haussier.

Resumen

Falta de seguimiento activo de los pacientes con cáncer en Chennai, India: implicaciones para las estimaciones de supervivencia basadas en la población

Objetivo Medir el sesgo de las estimaciones absolutas de la supervivencia de los enfermos de cáncer en ausencia de medidas de seguimiento activo de esos pacientes en los países en desarrollo.

Métodos El estudio abarcó todos los casos nuevos de los 10 cánceres más comunes y sus distintos subtipos, más todos los cánceres relacionados con el tabaco y no clasificados entre los 10 principales, que habían sido incluidos en el registro de cáncer basado en la población en Chennai, India, durante 1990-1999, y sometidos a seguimiento durante 2001. Los casos nuevos registrados se aparearon con los de la base de datos de mortalidad por todas las causas de la división de estadísticas vitales de la corporación municipal de Chennai, y los casos nuevos no apareados fueron sometidos luego a seguimiento activo para determinar su grado de supervivencia. La supervivencia absoluta se estimó mediante un método actuarial, aplicando diferentes supuestos respecto al estado de supervivencia (vivo/muerto) de los casos sometidos a seguimiento pasivo y activo.

Resultados Antes del seguimiento activo, el apareamiento entre casos osciló entre el 20% y el 66%, según la localización del tumor primario. El seguimiento activo de los casos nuevos no apareados reveló que entre un 15% y un 43% habían fallecido al final del periodo de seguimiento, y no se conocía el estado de supervivencia de un 4%-38% de los casos. Antes del seguimiento activo de los enfermos de cáncer, su supervivencia absoluta a los 5 años era según las estimaciones un 22%-47% superior a la determinada al aplicar los supuestos actuariales tradicionales a los casos perdidos para el seguimiento. Las estimaciones de supervivencia más bajas fueron las obtenidas al excluir de los análisis los casos perdidos para el seguimiento.

Conclusión En las condiciones reinantes en la India y en otros países en desarrollo, el seguimiento activo de los enfermos de cáncer es el método más fiable para estimar las tasas de supervivencia del cáncer. El simple seguimiento pasivo de los casos o la aplicación de los métodos habituales de estimación de la supervivencia tienden a ocasionar un sesgo por exceso.

ملخص

فقد المتابعة الفعّالة لمرضى السرطان في شّناي، الهند: تأثير تقديرات البُقيا السكانية

بعد ذلك متابعة فعّالة للتعرف على أوضاع الحالات التي لم تكن متوافقة مع قاعدة المعطيات، من حيث بقائها على قيد الحياة. وقدّر الباحثون المعدل المطلق للبقيا باستخدامهم طريقة حسابية (إكتوارية) وافتراسات مختلفة تتعلق بأوضاع البقيا (أحياء/أموات) للحالات التي طبق عليها المتابعة الفعّالة واللافاعلة.

الموجودات: قبل القيام بالمتابعة الفعّالة، تراوح التوافق بين الحالات بين 20% و66% ويعتمد ذلك على موقع الورم البدئي. فيما دلت المتابعة الفعّالة لوقوع الحالات غير المتوافقة أن ما يتراوح بين 15% و43% قد ماتوا قبل حلول نهاية فترة المتابعة فيما بقيت حالة البقيا لدى 4% إلى 38% منهم غير معروفة.

الهدف: قياس التحيز في تقديرات البُقيا (البقاء على قيد الحياة) المطلقة لدى مرضى السرطان عند فقد المتابعة الفعّالة لمرضاه في البلدان النامية.

الطريقة: شملت الدراسة جميع الحالات التي وقعت من أمهات السرطان العشرة الأكثر شيوعاً مع الأمهات الفرعية المتعلقة بها، والسرطانات المرتبطة بالتبغ والتي لم تصنّف ضمن تلك الأمهات العشرة، وسُجّلت في سجل سكاني للسرطان في شّناي، الهند، وذلك في خلال الفترة 1990 - 1999، وتمت متابعتها خلال عام 2001. وأجرى الباحثون مقارنة بين الحالات المسجلة أولاً وبين الحالات الموجودة في قاعدة معطيات لوفيات الناجمة عن جميع الأسباب في قسم الإحصاءات المدنية الحيوية في مؤسسة شّناي. ثم أجرى الباحثون

الاستنتاج: في ظل الظروف السائدة في الهند وفي البلدان النامية الأخرى، أدت المتابعة الفعّالة إلى الحصول على تقديرات لمعدل وفيات السرطان هي الأكثر موثوقية. أما المتابعة اللافاعلة للحالات أو تطبيق الطرق المعيارية لتقدير البُقيّة فتؤدي في الغالب إلى تحيُّزٍ يميل للزيادة.

وقبل المتابعة الفعّالة لمرضى السرطان، قدر الباحثون أن معدل البُقيّة المطلق لمدة 5 سنوات يزيد بمقدار يتراوح بين 22% و47% على ما سيكون عليه عند تطبيق الطرق التقليدية الحسابية الإكتوارية على الحالات التي فقدت من المتابعة. وقد حصل الباحثون على أقل قدر من تقديرات البُقيّة عند استبعاد الحالات التي فقدت من المتابعة أثناء التحليل.

References

- Parkin DM, Whelan SL, Ferlay J, Teppo L, Thomas DB. *Cancer incidence in five continents, Vol VIII* [IARC Scientific Publication No. 155]. Lyon: International Agency for Research on Cancer; 2002.
- National Cancer Registry Program. *Consolidated report of population based cancer registries*. New Delhi: Indian Council of Medical Research; 2006.
- Shanta V, Gajalakshmi CK, Swaminathan R, Ravichandran K, Vasanthi L. Cancer registration in Madras Metropolitan Tumour Registry, India. *Eur J Cancer* 1994;30A:974-8. PMID:7946594 doi:10.1016/0959-8049(94)90126-0
- Gajalakshmi CK, Shanta V, Swaminathan R. Cancer registration in India [IARC Suppl]. *Asian Pac J Cancer Prev* 2001;2:13-20.
- Swaminathan R, Shanta V, Rama R. Cancer registration, pattern and trend in India in the last two decades [Oncology update]. *Indian J Clin Pract* 2004;1:3-10.
- Nandakumar A, Gupta PC, Gangadharan P, Visweswara RN, Parkin DM. Geographic pathology revisited: development of an atlas of cancer in India. *Int J Cancer* 2005;116:740-54. PMID:15849747 doi:10.1002/ijc.21109
- Gajalakshmi CK, Shanta V, Rama R. Registration of cancer mortality data in a developing environment: Chennai (Madras) experience. *Cancer Causes Control* 1998;9:131-6. PMID:9578289 doi:10.1023/A:1008822008788
- Sankaranarayanan R, Black RJ, Swaminathan R, Parkin DM. An overview on cancer survival in developing countries. In: Sankaranarayanan R, Black RJ, Parkin DM, eds. *Cancer survival in developing countries* [IARC Scientific Publications No.145]. Lyon: International Agency for Research on Cancer; 1999.
- Cutler SJ, Ederer F. Maximum utilization of the life table method in analyzing survival. *J Chronic Dis* 1958;8:699-712. PMID:13598782 doi:10.1016/0021-9681(58)90126-7
- Berrino F, Sant M, Verdecchia, Capocaccia R, Hakulinen T, Esteve J. *Survival of cancer patients in Europe, The Eurocare Study* [IARC Scientific Publications No.132]. Lyon: International Agency for Research on Cancer; 1995.
- Wilson S, Prior P, Woodman CBJ. Use of cancer surveillance data for comparative analyses. *J Public Health Med* 1992;14:151-6. PMID:1515197
- Swaminathan R, Rama R, Shanta V. An appraisal of the active follow up system in the Hospital Cancer Registry, Cancer Institute (WIA), Chennai. *CRAB* 2003;X:23-9.
- Swaminathan R, Sankaranarayanan R, Hakama M, Shanta V. Effect of loss to follow up on population based cancer survival rates in developing countries. *Int J Cancer Suppl* 2002;13:172.
- Yeole BB, Jussawalla DJ, Sabnis SD, Sunny L. Survival from breast and cervical cancer in Mumbai (Bombay), India. In: Sankaranarayanan R, Black RJ, Parkin DM, eds. *Cancer survival in developing countries* [IARC Scientific Publication No. 145]. Lyon: International Agency for Research on Cancer; 1998. pp 79-87
- Vahdaninia M, Montazeri A. Breast cancer in Iran: a survival study. *Asian Pac J Cancer Prev* 2004;5:223-5. PMID:15244529
- Shanta V, Gajalakshmi CK, Swaminathan R. Survival from cancer in Chennai (Madras), India. In: Sankaranarayanan R, Black RJ and Parkin DM, eds. *Cancer survival in developing countries, Vol I* [IARC Scientific Publication No.145]. Lyon: International Agency for Research on Cancer; 1998. pp 89-100.
- Ganesh B. Effect of loss to follow up in estimating survival rates. *Acta Universitatis Tampereensis* (Tampere). Ser A 1995;440:1-66.
- Sriamporn S, Swaminathan R, Parkin DM, Kamsa-Ard S, Hakama M. Loss-adjusted survival of cervix cancer in Khon Kaen, Northeast Thailand. *Br J Cancer* 2004;91:106-10. PMID:15199396 doi:10.1038/sj.bjc.6601959

Childhood cancers in Chennai, India, 1990–2001: Incidence and survival

Rajaraman Swaminathan^{1*}, Ranganathan Rama¹ and Viswanathan Shanta²

¹Division of Epidemiology and Cancer Registry, Cancer Institute (WIA), Chennai, Tamil Nadu, India

²Department of Medical Oncology, Cancer Institute (WIA), Chennai, Tamil Nadu, India

Childhood cancers (age at diagnosis: 0–14 years) comprise a variety of malignancies, with incidence varying worldwide by age, sex, ethnicity and geography, that provide insights into cancer etiology. A total of 1,334 childhood cancers registered in population-based cancer registry, Chennai, India, during 1990–2001 and categorized by International Classification of Childhood Cancer norms formed the study material. Cases included for survival analysis were 1,274 (95.5%). Absolute survival was calculated by actuarial method. Cox proportional hazard model was used to elicit the prognostic factors for survival. The age-standardized rates for all childhood cancers together were 127 per million boys and 88 per million girls. A decreasing trend in incidence rates with increasing 5-year age groups was observed in both sexes. The top 5 childhood cancers were the same among boys and girls: leukemias, lymphomas, central nervous system neoplasms, retinoblastomas and renal tumors. The highest 5-year absolute survival was observed in Hodgkin's disease (65%) followed by Wilm's tumor (64%), retinoblastomas (48%), non-Hodgkin's lymphomas (47%), osteosarcomas (44%), acute lymphoid leukemia and astrocytoma (39%). Multifactorial analysis of age at diagnosis and sex showed no differences in the risk of dying for all childhood cancers. Completeness of treatment and type of hospital combination emerged as a prognostic factor for survival for all childhood cancers together ($p < 0.001$), acute lymphoid leukemia ($p < 0.001$) and non-Hodgkin's lymphoma ($p = 0.04$). A Childhood Cancer Registry with high-resolution data collection is advocated for in-depth analysis of variation in incidence and survival.

© 2008 Wiley-Liss, Inc.

Key words: population-based cancer registry; childhood cancer; incidence; survival; childhood cancer registry

The network of cancer registries called National Cancer Registry Program of Indian Council of Medical Research, Government of India, has been in existence since 1981.¹ The Madras Metropolitan Tumor Registry (MMTR), a population-based cancer registry (PBCR), based at the Cancer Institute (WIA), Chennai, collects data on all incident cancer cases occurring in Chennai city and publishes reports on incidence and cancer survival to serve as reference material for oncologists involved in clinical work and research.^{2–4} Childhood cancers (age at diagnosis 0–14 years) comprise a variety of malignancies, with incidence varying worldwide by age, sex, ethnicity and geography, that provide insights into cancer etiology.^{5,6} Publications on childhood cancer incidence and, especially, survival are sparse from India.^{7,8} This paper analyses the incidence of and survival from childhood malignancies in Chennai during the period 1990–2001.

Material and methods

Cancer is not a notifiable disease in India. Hence, registration of cases is done by active method.² The MMTR covers an entire urban population (all ages) estimated at 4.1 millions in 1996 in the ratio of 960 females to 1,000 males. The corresponding childhood population figures were 1,067,425 in the ratio of 964 girls to 1,000 boys. Childhood cancers are categorized by International Classification of Childhood Cancer norms⁵ using ICD-O II edition codes.⁹ A total of 1,334 childhood cancers registered in MMTR during 1990–2001 formed the study material. Comparative measures of incidence like age-specific rate and age-standardized rate per million and cumulative risk expressed as ratio at risk were calculated to describe the data on cancer incidence for both boys and girls.⁶

Data on mortality were routinely collected from the vital statistics division.¹⁰ Deaths occurring in Chennai were routinely matched with the incident cancer database. The incident cancer cases that were not matched with the mortality database were subjected to follow-up by active methods: repeated scrutiny of medical records, postal/telephone enquiry and/or house visits. The closing date of follow-up was December 31, 2003. Cases registered based on a death certificate only (1.8%) and those with no information on follow-up apart from the incidence date (2.7%) were excluded from survival analysis. Thus, 1,274 cases were included for survival analysis (95.5%). Data on treatment was routinely abstracted 6 months to 1 year from diagnosis. Complete treatment signified those who had received at least 1 full first course of cancer-directed treatment independent of the type of treatment. Incomplete treatment referred to starting of cancer treatment, but failure to complete it at the time of abstracting this information. Type of hospital was broadly classified into 2 categories in this study: (i) public hospitals in government sector and (ii) hospitals including comprehensive cancer centers in nongovernment sector. The survival time was calculated as the duration between the incidence date and the date of death or date of loss to follow-up or closing date, whichever was earlier. Absolute survival was calculated by actuarial method.¹¹ Log rank test was used to compare the survival curves.¹² Cox proportional hazard model was used to elicit the main effects in multifactorial analysis of prognostic factors for survival.¹³

Results

Data quality indices for the study of incidence and survival from childhood cancers are given in Table I. Among the 1,334 childhood cancers studied for computing measures of incidence, histologically verified cancer diagnosis comprised 1,272 (95%), and cases included for survival study constituted 1,274 (95%). Complete follow-up achieved with definite knowledge of alive/dead status was 85% at 5 years from diagnosis and 74% at 10 years from diagnosis.

Descriptive statistics on childhood cancer morphology groups by 5-year age groups and sex are given in Table II. A preponderance of boys in the incidence was forthcoming for majority of morphology and 5-year age groups. The age-specific incidence rates per million of all childhood cancers together in the age groups 0–4 years, 5–9 years and 10–14 years were 143, 128 and 104 among boys and 112, 76 and 68 among girls, respectively. A decreasing trend in incidence rates with increasing 5-year age groups was observed, with the peak incidence occurring at 0–4 years of age in both sexes. The age-standardized rates per million for all childhood cancers together were 127 for boys and 88 for girls. The ratio at risk of acquiring cancer in the pediatric age group was 1 in 534 among boys and 1 in 780 among girls. The top 5 morphology groups in childhood cancers based on number of

Grant sponsors: Indian Council of Medical Research, New Delhi, India; International Agency for Research on Cancer, Lyon, France.

*Correspondence to: Division of Epidemiology and Cancer Registry, Cancer Institute (WIA), 38 Sardar Patel Road, Chennai 600036, India. Fax: +91-44-24912085. E-mail: iarcsurvival@yahoo.co.uk

Received 25 September 2007; Accepted after revision 17 December 2007

DOI 10.1002/ijc.23428

Published online 6 March 2008 in Wiley InterScience (www.interscience.wiley.com).

cases were the same among boys and girls: leukemias, lymphomas, central nervous system neoplasms, retinoblastomas and renal tumors. Acute lymphoid leukemia (76%) was the commonest among leukemias; Hodgkin's disease accounted for half of all lymphomas together among boys and lesser among girls. Retinoblastomas were evenly distributed among both sexes.

Absolute survival from all incident childhood cancers in Chennai, both treated and untreated ones, at one, three, 5 and 10 years from diagnosis were 65%, 46%, 40% and 35%, respectively. The highest 5-year absolute survival among morphology groups and subgroups was observed in Hodgkin's disease (65%) followed by

Wilm's tumor (64%), retinoblastomas (48%), non-Hodgkin's lymphomas (47%), osteosarcomas (44%), acute lymphoid leukemia and astrocytoma (39%). The poorest survival was seen in hepatic tumors (11%) preceded by Ewing's sarcoma (23%; Table III). However, the differences in survival between 5-year age groups in both sexes were not statistically significant for any cancer.

The 5-year absolute survival (%) and multifactorial analysis of prognostic factors for survival from all childhood cancers together and selected childhood cancers in Chennai are given in Table IV. The factors analyzed were as follows: 5-year age group at diagnosis (3-levels: 0-4, 5-9 and 10-14 years), sex (2-levels: boys and girls) and treatment status+type of hospital (4-levels): (i) Completed treatment in nongovernment hospitals, (ii) completed treatment in public hospitals in government sector, (iii) registered but received incomplete or no treatment in public hospitals in government sector and (iv) registered but received incomplete or no treatment in nongovernment hospitals). The differences in survival by treatment status+type of hospital were statistically significant for all childhood cancers together, acute lymphoid leukemia and non-Hodgkin's lymphoma. In multifactorial analysis of prognostic factors, the combination of completeness of treatment and type of hospital emerged as a significant prognostic factor for survival from all childhood cancers together ($p < 0.001$). A 2- to 5-fold increased risk of dying among acute lymphoid leukemia cases ($p < 0.001$) was observed for those who received complete treatment in public hospitals in government sector and for those with incomplete or no treatment in any hospital compared to receiving complete treatment in nongovernment hospitals. The risk of dying among non-Hodgkin's lymphoma cases was 4-fold higher ($p = 0.01$) for those not treated or received incomplete treatment in public hospitals in government sector compared to those receiving complete treatment in nongovernment hospitals. No statistically significant differences in the risk estimates were forthcoming

TABLE I - DATA QUALITY INDICES: INCIDENCE OF AND SURVIVAL FROM CHILDHOOD CANCERS, CHENNAI, 1990-2001

	Number	%
Data quality indices		
<i>Incidence data</i>		
Total cases	1,334	-
Histological verification	1,272	95.4
<i>Excluded from survival analysis</i>		
Death certificate only cases	24	1.8
Cases with no follow-up	36	2.7
Cases included for survival analysis	1,274	95.5
<i>Follow-up data of included cases</i>		
Cases with complete follow-up at 10 years from diagnosis	942	73.9
Alive	230	18.0
Dead	712	55.9
Partial or lost to follow-up	332	26.1
Know alive <1 year from diagnosis	83	6.5
1-3 years from diagnosis	74	5.8
3-5 years from diagnosis	30	2.4
5-10 years from diagnosis	145	11.4

TABLE II - DESCRIPTIVE STATISTICS ON CHILDHOOD CANCER INCIDENCE BY MAJOR MORPHOLOGY GROUPS AND SEX: CHENNAI, 1990-2001

Childhood cancer classification: morphology groups	Average annual rates per million and risk											
	All cases		Boys (B)					Girls (G)				
			Age-specific rates			ASR	Ratio at risk	Age-specific rates			ASR	Ratio at risk
Number of cases	G:B	0-4	5-9	10-14	0-4			5-9	10-14			
Leukemia	463	1:1.7	50.8	50.0	34.5	45.8	1:1,478	41.3	24.3	18.8	29.3	1:2,369
Lymphoid leukemia	353	1:1.8	40.1	40.0	25.3	35.7	1:1,899	34.3	17.3	11.4	22.2	1:3,175
Acute lymphoid leukemia	351	1:1.8	40.1	39.5	24.8	35.5	1:1,915	34.3	17.3	11.4	22.2	1:3,175
Acute nonlymphoid leukemia	63	1:1.3	4.6	7.3	4.6	5.5	1:12,102	3.2	4.2	5.3	4.1	1:15,772
Acute myeloid leukemia	60	1:1.2	4.1	6.4	4.6	5.0	1:13,242	3.2	4.2	5.3	4.1	1:15,772
Chronic myeloid leukemia	7	1:0.8	0.5	0.5	0.4	0.5	1:143,970	0.5	0.5	0.9	0.6	1:106,386
Lymphomas and reticuloendothelial neoplasm	271	1:2.7	14.9	42.3	32.0	28.7	1:2,244	4.8	16.8	12.3	10.9	1:5,897
Hodgkin's disease	126	1:3.5	5.6	24.1	14.3	14.1	1:4,540	1.6	6.5	4.8	4.1	1:15,419
Non-Hodgkin lymphoma	112	1:2.4	7.2	13.6	14.7	11.5	1:5,624	2.7	8.4	4.4	5.0	1:12,925
CNS and misc intracranial and intraspinal neoplasms	142	1:1.4	12.3	15.0	10.5	12.7	1:5,284	8.0	14.5	6.1	9.6	1:6,978
Astrocytoma	70	1:1.1	6.7	5.0	5.5	5.8	1:11,662	4.8	7.5	3.5	5.3	1:12,653
Sympathetic nervous system tumors	56	1:0.8	7.7	2.3	2.1	4.3	1:16,554	10.7	4.2	0.9	5.8	1:12,655
Neuroblastoma and ganglioneuroblastoma	54	1:0.8	7.7	2.3	1.7	4.2	1:17,152	10.7	3.7	0.9	5.6	1:13,041
Retinoblastoma	82	1:1.0	15.9	4.1	0.4	7.6	1:9,789	16.1	4.7	0.4	7.9	1:9,437
Renal tumors	72	1:1.1	14.4	3.6	0.4	6.9	1:10,848	14.5	3.3	0.4	6.8	1:10,999
Wilm's tumor	65	1:1.0	13.4	3.2	0.0	6.2	1:12,096	13.9	2.8	0.0	6.3	1:11,946
Hepatic tumors	20	1:0.8	4.1	0.5	0.0	1.7	1:43,830	5.4	0.5	0.0	2.2	1:34,315
Malignant bone tumors	52	1:1.3	1.0	4.1	7.6	3.9	1:15,752	0.0	1.4	8.8	3.0	1:19,671
Osteosarcoma	26	1:1.2	0.5	0.9	4.6	1.8	1:33,033	0.0	1.4	3.9	1.6	1:37,409
Ewing's sarcoma	20	1:1.2	0.5	2.3	2.1	1.5	1:40,887	0.0	0.0	3.9	1.1	1:50,703
Soft tissue sarcoma	62	1:1.6	8.2	3.2	6.3	6.0	1:11,290	4.8	2.8	3.9	3.9	1:17,282
Rhabdomyosarcoma	43	1:1.4	6.2	2.7	2.9	4.1	1:16,896	4.8	1.9	2.2	3.1	1:22,509
Germ cell, trophoblastic and other gonadal neoplasms	45	1:0.6	8.2	0.5	0.0	3.3	1:23,064	3.8	0.9	8.3	4.2	1:15,368
Other carcinoma and unspecified malignant neoplasm	69	1:1.3	5.1	2.7	9.7	5.7	1:11,398	2.7	2.8	8.3	4.4	1:14,481
<i>All childhood cancers together</i>	<i>1,334</i>	<i>1:1.5</i>	<i>142.8</i>	<i>128.2</i>	<i>103.6</i>	<i>126.7</i>	<i>1:534</i>	<i>112.1</i>	<i>76.2</i>	<i>68.4</i>	<i>87.8</i>	<i>1:780</i>

ASR, age-standardized rate.

TABLE III – ABSOLUTE SURVIVAL FROM ALL AND MAJOR CHILDHOOD CANCERS CHENNAI, 1990–2001, CASES FOLLOWED THROUGH 2003

	No. of cases	Absolute survival			
		1 year	3 year	5 year	10 year
Leukemia	446	58.6	40.5	36.3	29.8
Lymphoid leukemia	345	59.4	43.5	38.7	35.4
Acute lymphoid leukemia	343	59.4	43.5	38.7	35.3
Acute nonlymphoid leukemia	60	55.2	31.0	31.0	16.1
Acute myeloid leukemia	58	53.6	30.3	30.3	15.7
Chronic myeloid leukemia	7	71.4	25.7	25.7	–
Lymphomas and reticuloendothelial neoplasm	270	72.4	60.7	55.3	49.3
Hodgkin's disease	126	83.5	72.1	65.0	65.0
Non-Hodgkin lymphoma	111	61.7	49.3	46.6	35.7
CNS and misc intracranial and intraspinal neoplasms	134	56.6	30.2	26.8	25.0
Astrocytoma	68	66.9	43.6	38.7	34.9
Sympathetic nervous system tumors	55	65.1	36.3	25.7	25.7
Neuroblastoma and ganglioneuroblastoma	64	45.7	34.6	36.9	26.9
Retinoblastoma	72	87.0	59.6	48.1	45.0
Renal tumors	66	74.4	61.0	58.0	46.9
Wilm's tumor	61	77.6	67.6	64.2	52.0
Hepatic tumors	19	21.1	15.8	10.5	10.5
Malignant bone tumors	46	64.8	44.1	30.6	30.6
Osteosarcoma	22	58.1	43.6	43.6	43.6
Ewing's sarcoma	19	73.7	46.9	23.4	–
Soft tissue sarcoma	61	66.7	45.4	36.3	36.3
Rhabdomyosarcoma	42	68.3	47.1	36.4	36.4
Germ cell, trophoblastic and other gonadal neoplasms	45	71.4	54.5	38.0	38.0
Other carcinoma and specified malignant neoplasm	60	70.0	47.6	35.1	–
All childhood cancers together	1,274	64.9	46.4	40.0	35.3

TABLE IV – FIVE-YEAR ABSOLUTE SURVIVAL AND MULTIFACTORIAL ANALYSIS OF PROGNOSTIC FACTORS FOR SURVIVAL FROM CHILDHOOD CANCERS, CHENNAI: 1990–2001

Factors	All childhood cancers together			Acute lymphoid leukemia			Non-Hodgkin's lymphoma		
	n	5-yr survival	HR ¹	n	5-yr survival	HR ²	n	5-yr survival	HR ²
Age group (in years)									
00–04	460	40.0	1.0 ³	139	43.3	1.0 ³	19	52.6	1.0 ³
05–09	431	43.5	0.9 (0.8–1.1) ⁴	120	37.2	1.2 (0.8–1.6)	48	46.0	1.3 (0.6–2.8)
10–14	383	36.1	1.0 (0.9–1.2)	84	33.4	1.3 (0.9–1.8)	44	44.1	1.0 (0.4–2.2)
Sex									
Male	772	39.2	1.0 ³	218	37.6	1.0 ³	78	44.1	1.0 ³
Female	502	41.4	1.9 (0.8–1.1)	125	41.1	1.0 (0.8–1.3)	33	51.2	0.7 (0.4–1.3)
Treatment+type of hospital									
Complete+nongovernment	284	54.3	1.0 ³	65	60.0	1.0 ³	20	71.4	1.0 ³
Complete+government	797	40.1	1.4 (1.2–1.8) ⁵	254	37.4	1.7 (1.1–2.5) ⁵	72	44.5	1.9 (0.8–4.7)
Nil/incomplete+government	125	25.1	2.6 (1.9–3.4) ⁵	12	8.3	5.2 (2.5–10.6) ⁵	14	21.4	3.6 (1.3–9.9) ⁵
Nil/incomplete+nongovernment	68	12.1	3.2 (2.3–4.4) ⁵	12	0.0	5.0 (2.5–10.1) ⁵	5	60.0	1.3 (0.2–6.5)

HR, hazard ratio; CI, confidence interval.

¹Each factor above has been adjusted for the others in the model+morphology group. ²Each factor above has been adjusted for the others in the model. ³Reference category. ⁴Values given in parentheses indicate 95% CIs. ⁵ $p < 0.05$.

between treatment status+types of hospitals for the rest of childhood cancers. Age at diagnosis and sex did not emerge as significant risk factors for any of the childhood cancers studied.

Discussion

The MMTR is one of the oldest cancer registries in the network of PBCRs in India. The data on cancer incidence have been published in several volumes of the International Agency for Research on Cancer Scientific Publication series on "Cancer Incidence in 5 Continents," without any reservation that stands testimony to the data quality.⁶ Completeness of coverage of the Chennai registry is estimated as 96%.¹⁴ This registry was the first to initiate active follow-up of cancer cases in a PBCR in India, and the loss to follow-up encountered in this study is the least among all PBCRs in India.¹⁵ The population-based survival estimation is unbiased by selection, as it pertains to all cancers that are treated as well as not treated in the registry area, thus portraying an average outcome from the disease in Chennai city. Also, exclusion of cases from survival analysis has been very minimal in this study. Loss-adjusted survival rates¹⁶ were estimated to confirm that the con-

ventional estimates were not biased because of the frequent losses to follow-up (26% at 10 years from diagnosis). Hence, the data on childhood cancer incidence and survival from Chennai are fairly reliable for any population-based comparison with other registries.

The age-standardized rate of all childhood cancers and both sexes together in Chennai was 108 per one million children, which is the highest among urban registries in India.¹ The trend of age-standardized rates of all childhood cancers together during the two decades between 1982 and 2001 in Chennai showed an average annual increase of 2% among boys and 1% among girls. The variation in the increase in cancer incidence among boys and girls in the last two decades in Chennai may have been due to the decreasing trend in the background childhood population at risk, especially girls, in Chennai.^{17–19} The increase in the childhood cancer incidence in general might be consequent to the advent of newer facilities for diagnosis of a majority of childhood cancers in Chennai, which also ensures completeness of cancer registration. However, noticeable increase was observed only in lymphoid leukemia, brain, renal and soft tissue tumors. A decrease was forthcoming in non-Hodgkin's lymphoma and retinoblastoma, which may indicate that changes also in etiological exposures had taken place.

TABLE V – COMPARISON OF 5-YEAR SURVIVAL FROM CHILDHOOD CANCERS IN CHENNAI, BANGALORE,⁷ ACCIS²¹ AND US-SEER¹⁸ STUDIES

ICCC morphology group	ICCC code	Chennai	Bangalore	ACCIS	US-SEER
		1990–2001	1982–1989	1988–1997	1985–1989
Lymphoid leukemia	Ia	39	36	79	77
Acute nonlymphoid leukemia	Ib	31	10	49	34
Hodgkin's disease	IIa	65	73	93 ¹	90
Non-Hodgkin's lymphoma	IIb	47	33	85	73
Astrocytoma	IIIb	39	41	75	72
Neuroblastoma and ganglioneuroblastoma	IVa	37	29	59	57
Retinoblastoma	V	48	73	93	87
Wilm's Tumour	VIa	64	28	83	90
Osteosarcoma	VIIIa	44	45	59	68
Ewing's sarcoma	VIIIc	23	–	62	61
Rhabdomyosarcoma	IXa	36	14	63	60

ICCC, international childhood cancer classification; ACCIS, automated childhood cancer information system.

¹Includes cases age 1–14 only.

The incidence among boys is not very different between Chennai and most developed countries, but the incidence among girls in Chennai is significantly less than in most developed countries.⁶ This is observed also in other registries in India and other developing countries.^{1,6} It has been suggested that such elevated sex ratios among childhood cancers might reflect the socioeconomic level of the society in which girls are more likely to be underdiagnosed than boys, though the health care is free at the point of delivery.²⁰ But, even in a more developed society with minimal socioeconomic inequalities like Singapore, an elevated sex ratio in the childhood cancer incidence among Singapore Indians (49 girls to 100 boys) was observed compared to 93:100 among Singapore Malay and 78:100 among Singapore Chinese.⁶ The Singapore Indian population is not ethnically any different from the population of Tamil Nadu in India, whose capital city is Chennai. Thus, the underlying nature and etiology of the disease may also be important factors for this differential sex pattern observed in childhood cancer incidence in Chennai. The age-standardized rate of Hodgkin's disease among boys in Chennai (15 per million) was higher than developed countries.^{1,6} The first peak in the age-specific incidence curve for Hodgkin's disease was observed in the childhood in Chennai, and it occurred much later in other developed countries.⁶ This is consistent either with different etiology or with different number of oncogenes or causes that activate them. Retinoblastomas in Chennai and Delhi are either on par or more than developed countries.^{5,8}

Table V gives the comparison in absolute survival of childhood cancers between Chennai, Bangalore, US-SEER and ACCIS studies.^{7,21–23} Absolute survival is lower by one-half in Chennai for lymphoid leukemia, astrocytoma, retinoblastoma and Ewing's sarcoma compared to US-SEER and/or ACCIS studies. Absolute differences in survival between Chennai and US-SEER and/or ACCIS studies were more than 20 units for Hodgkin's disease, non-Hodgkin's lymphoma, Wilm's tumor and osteosarcoma, with less survival in Chennai. Survival was comparable in all the studies for acute nonlymphoid leukemia, with the exception of Bangalore where it was the least. Comparison of absolute survival between Chennai and Bangalore revealed only minimal differences for most cancers: survival was notably higher in Chennai for acute nonlymphoid leukemia, Wilm's tumor and rhabdomyosarcoma, and less for retinoblastoma than Bangalore. This wide variation in survival from childhood cancers between India and difference, poorer than in developed countries, are likely due to conditions for treatment and to the other differences in socioeconomic status, cancer awareness and stage of disease. Therefore, it is not possible to draw any conclusions on the relationship on etiology and survival.

In Chennai, the infrastructure for treatment of childhood cancers in terms of development and accessibility are available to all sections of the society. The Cancer Institute (WIA) is a nongovernment institution and the first Regional Cancer Centre in the region with all state-of-art diagnostic and treatment facilities for

cancer, wherein more than two-third of its cases received highly subsidized or free treatment. In the government sector, there are (i) a specialty teaching hospital dedicated to children, with the required infrastructure for management of all childhood cancers at free of cost and (ii) 4 other teaching hospitals catering to all ages with facilities for radiotherapy, surgery and chemotherapy, also at free of cost. In the private sector, there is a specialty hospital dedicated for children with cancer consultants, 6 other hospitals with radiotherapy facility catering to cancer patients of all ages, and many others with cancer treatment facilities excluding radiotherapy. However, the burden of cancer cases catered to by these hospitals is quite heavy and not restricted only to those cases that are residents of Chennai city. Also, the pattern of care including treatment protocols followed for common childhood cancers in the different hospitals in Chennai were not uniform. It is, however, not clear how these factors affect the cancer survival by treatment status and type of hospitals. But it seems that while the survival in public hospitals in government sector was less than in hospitals in nongovernment sector for patients with complete treatment, the reverse was true for the incomplete treatment. This probably indicates that economic constraints in living conditions affected the completeness of treatment more than the cost of hospitalization and treatment itself.

With this background, the multifactorial analysis of prognostic factors for survival from all childhood cancers together and the major childhood cancers in Chennai did not show any statistically significant differences by age at diagnosis and sex, as observed in similar studies from developed countries.^{24,25} The increasing risk of death from acute lymphoid leukemia by increasing age group signifying an inverse relationship between survival and age at diagnosis was observed in this study as well as EURO CARE studies²⁴ but not statistically significant. In this study, significant changes in the risk of dying were observed for all childhood cancers together when treatment status and type of hospital were combined and adjusted for the morphology group. However, the lack of information on the following is conspicuous: (i) stage of disease, (ii) socioeconomic status, (iii) patient compliance to first course of treatment for the entire duration which is crucial, (iv) type of treatment or protocols followed and (v) supportive care given to specific cancer types or individual patients. Hence, the observation on the statistically significant risk estimates by type of hospitals should be interpreted with caution.

In India and developing countries in general, the etiology of many childhood cancers is likely to be different from that in more developed countries. When all childhood cancers occurring in a region are considered, treatment still remains sometimes incomplete, and survival in general is less in India than in more developed countries. However, for children completing treatment in the comprehensive cancer centers, the survival approaches that in Europe or USA. A Childhood Cancer Registry is indicated, with emphasis on collection of high-resolution data for explaining the variation.

Acknowledgements

We are grateful to the cancer registrars of MMTR for diligent data collection and other registry staff for data processing and are

indebted to all the medical institutions and the vital statistics division in and around the city of Chennai for providing data to the population-based cancer registry, without which this study would not have been a reality.

References

1. National Cancer Registry Programme. Consolidated report of population-based cancer registries 1990–1996: incidence and distribution of cancer. New Delhi, India: Indian Council of Medical Research, 2001.
2. Shanta V, Gajalakshmi CK, Swaminathan R, Ravichandran K, Vasanthi L. Cancer registration in Madras Metropolitan Tumour Registry. *Eur J Cancer* 1994;30:974–8.
3. Swaminathan R, Shanta V, Rama R. Cancer registration, pattern and trend in India in the last two decades. In: Shanta V, Rajkumar T, eds. *Postgraduate Oncology Update – Indian Journal of Clinical Practice*. New Delhi: IJCP Group of Publications, 2004. pp 3–10.
4. Shanta V, Gajalakshmi CK, Swaminathan R. India: Madras Metropolitan Tumour Registry, 1982–92. In: Parkin DM, Kramarova E, Draper GJ, Masuyer E, Michaelis J, Neglia S, Qureshi S, Stiller CA, eds. *International incidence of childhood cancer, Vol. 2*. Lyon: International Agency for Research on Cancer, 1998. IARC Scientific Publication No. 144, pp 169–72.
5. Kramarova E, Stiller CA. International classification of childhood cancer (ICCC). In: Parkin DM, Kramarova E, Draper GJ, Masuyer E, Michaelis J, Neglia S, Qureshi S, Stiller CA, eds. *International incidence of childhood cancer, Vol. 2*. Lyon: International Agency for Research on Cancer, 1998. IARC Scientific Publication No. 144. pp 15–9.
6. Parkin DM, Whelan SL, Ferlay J, Teppo L, Thomas DB, eds. *Cancer incidence in five continents, Vol. 8*. Lyon: International Agency for Research on Cancer, 2002. IARC Scientific Publication No. 155.
7. Nandakumar A, Anantha N, Kumaraswami A, Appaji L, Mukherjee G, Venugopal T, Dhar M. Descriptive epidemiology of childhood cancers in Bangalore, India. *Cancer Causes Control* 1996;7:405–10.
8. Tyagi BB, Manoharan N, Raina V. Childhood cancer incidence in Delhi, 1996–2000. *Indian J Med Paed Oncol* 2006;27:13–8.
9. WHO. *International Classification of Diseases for Oncology (ICD-O)*, 2nd edn. Geneva: World Health Organization, 1990.
10. Gajalakshmi CK, Shanta V, Rama R. Registration of cancer mortality data in developing environment: Chennai (Madras) experience. *Cancer Causes Control* 1998;9:131–36.
11. Cutler SJ, Ederer F. Maximum utilization of life table method in analyzing survival. *J Chronic Dis* 1958;8:699–712.
12. Mantel N. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep* 1966;50:163–70.
13. Cox DR. Regression models and life tables. *J R Statist Soc (B)* 1972;34:187–220.
14. Gajalakshmi V, Swaminathan R, Shanta V. An independent survey to assess completeness of registration: Population-Based Cancer Registry, Chennai, India. *Asian Pac J Cancer Prev* 2001;2:179–83.
15. Sankaranarayanan R, Black RJ, Swaminathan R, Parkin DM. An overview on cancer survival in developing countries. In: Sankaranarayanan R, Black RJ, Parkin DM, eds. *IARC monograph on cancer survival in developing countries*. Lyon: International Agency for Research on Cancer, 1998. IARC Scientific Publication No. 145. pp 19–26.
16. Sriamporn S, Swaminathan R, Parkin DM, Kamsa-Ard S, Hakama M. Loss-adjusted survival of cervix cancer in Khon Kaen, Northeast Thailand. *Br J Cancer* 2004;91:106–10.
17. Census of India 1981. Social and cultural tables, Series-20, Tamil Nadu: Part IV-A, Table C-6. Chennai: Directorate of Census Operations, 1987. 329 p.
18. Census of India 1991. Social and cultural tables, Series-23, Tamil Nadu: Part IV-A, Table C-6. Chennai: Directorate of Census Operations, 1997. 534 p.
19. Census of India 2001. Final population totals, Series-34, Tamil Nadu. Chennai: Directorate of Census Operations, 2005.
20. Pearce MS, Parker L. Childhood cancer registrations in the developing world: still more boys than girls. *Int J Cancer* 2001;91:402–406.
21. Steliarova-Foucher E, Coebergh JW, Kaatsch P, Pritchard-Jones K, Stiller C, eds. *Cancer in children and adolescents in Europe*. *Eur J Cancer* 2006;42:1913–2190.
22. Gatta G, Capocaccia R, Coleman MP, Ries LA, Berrino F. Childhood cancer survival in Europe and the United States. *Cancer* 2002; 95:1767–72.
23. Ries LAG, Smith MA, Gurney JG, Linet M, Tamra T, Young JL, Bunin GR, eds. *Cancer incidence and survival among children and adolescents: United States SEER Program 1975–1995*. Bethesda: National Cancer Institute SEER Program, 1999. NIH Pub. No. 99-4649.
24. Coebergh JWW, Pastore G, Gatta G, Corazziari I, Kamps W; EURO-CARE Working Group. Variation in survival of European children with acute lymphoblastic leukemia, diagnosed in 1978–1992: the EURO-CARE study. *Eur J Cancer* 2001;37:687–94.
25. Pastore G, Magnani C, Verdecchia A, Pession A, Viscomi S, Coebergh JWW; EURO-CARE Working Group. Survival of childhood lymphomas in Europe, 1978–1992: a report from the EURO-CARE study. *Eur J Cancer* 2001;37:703–10.

Loss-adjusted survival of cervix cancer in Khon Kaen, Northeast Thailand

S Sriamporn^{*1}, R Swaminathan², DM Parkin³, S Kamsa-ard⁴ and M Hakama⁵

¹Department of Epidemiology, Faculty of Public Health, Khon Kaen University, Khon Kaen 40002, Thailand; ²Cancer Institute (WIA), Chennai, India; ³International Agency for Research on Cancer, Lyon, France; ⁴Cancer Unit, Srinagarind Hospital, Khon Kaen University, Thailand; ⁵Tampere School of Public Health, University of Tampere, Finland

For incident cancers of the cervix uteri (601 cases) registered in the population-based cancer registry of Khon Kaen province, Northeast Thailand, in 1985–1990 loss-adjusted survival probabilities were estimated by a logistic regression model with four prognostic factors (age at diagnosis, stage of disease, place of residence and treatment), and compared with observed survival, estimated by the actuarial method. All patients were followed up for a minimum of 5 years, using both passive and active methods. In all, 27.6% of patients were lost to follow-up within 5 years of the index date. The overall observed survival at 5 years was 56.8% and loss-adjusted survival was 54.7%. The difference between the loss-adjusted and observed survival at 5 years was small: 2.1% overall, varying between 0.8 and 3.5 percent units for any prognostic group. The assumption of independence of loss to follow-up and death in the calculation of survival by the actuarial method in this, and probably in other, population-based series, is reasonable and leads to no material bias in the estimates.

British Journal of Cancer (2004) 91, 106–110. doi:10.1038/sj.bjc.6601959 www.bjcancer.com

Published online 15 June 2004

© 2004 Cancer Research UK

Keywords: cervix cancer; survival; loss to follow-up

Population level survival is usually estimated by the life-table method, in which cumulative probability of survival is calculated at successive annual intervals after diagnosis (Cutler & Ederer, 1958; Ederer *et al*, 1961). Information from all cases is used, including cases whose follow-up ends due to closure of the study, and those lost to follow-up before closure. Survival estimates may be biased if the proportion of cases lost to follow-up is substantial (as in many developing countries, where health information systems are not well developed), and if the loss to follow-up is correlated with the probability of death (prognosis) of the patient after he or she was lost.

Prognostic factors that may also predict loss to follow-up are related to the clinical characteristics of the disease, the patient and the social environment. For example, recurrence or relapse of the disease and serious comorbidity are prognostic factors that may cause the patient to move away (for treatment, or terminal care), making them impossible to trace. Social status influences the probability of survival from cancer (Kogevinas and Porta, 1997) and may also affect the ability to follow-up of a subject.

Information on the association between prognostic factors and loss to follow-up can be used to reduce the bias in estimates of survival (Ganesh, 1995; Mathew, 1996). In this paper, we calculate the absolute survival of cases of cancer of the cervix recorded by a population-based cancer registry in Thailand, using the actuarial method, and examine the effect of adjustment for differential loss

to follow-up within subgroups of patients at different risk of death from the disease ('loss adjustment').

SUBJECTS AND METHODS

A total of 630 invasive incident cancers of the cervix were registered during 1985–1990 in the population-based cancer registry covering the province of Khon Kaen. Of these, 29 (4.6%) were registered on the basis of a death certificate only, and were excluded from the survival analysis. For the remaining 601 cases, data on age at diagnosis, area (district) of residence, date of incidence, topography, morphology, stage of disease, treatment (whether treatment by surgery, radiation or chemotherapy was recorded in the patient file), date and vital status (alive or dead) at last contact were abstracted from the registry database.

Patients were followed up until death, or date of loss to follow-up, or 31 December 1995 (closing date). Therefore, the potential length of follow-up was 5–10 years. The registry used both passive and active measures to establish the vital status (alive/dead) of cancer patients.

Passive follow-up

All death certificates (with a mention of any cancer (ICD-9: 140–208) as underlying or contributing cause of death) were obtained from the Provincial Health Department. The death certificates were linked to the cancer registry database at annual intervals (using national ID number, name, date of birth and address) and the date of death updated for matching cases.

*Correspondence: Dr S Sriamporn; E-mail: supanee@kku.ac.th

Received 22 December 2003; revised 23 April 2004; accepted 23 April 2004; published online 15 June 2004

Active follow-up

For the remaining unmatched cases, information on follow-up was collected by visiting the various hospitals to scrutinise case records, and by making enquiries of treating physicians and general practitioners. Annual follow-up on the anniversary of the date of incidence was attempted for presumed survivors by sending a reply-paid postcard inquiring about the current status of the patient. If no reply was received, a second postcard was sent to the headman of the village requesting the same information. House visits were also performed wherever feasible.

Analytical methods

Actuarial survival The estimation of survival probability for each year was carried out by the actuarial method. The index date of this study was the date of incidence. The duration of survival for each case was calculated as the time elapsed from the index date to the date of death or the last date of follow-up or closing date, whichever was earlier. Cumulative absolute survival (Cutler and Ederer, 1958) was estimated using the SURV3 analysis programme (Dickman *et al*, 2002).

Loss-adjusted survival The method proposed by Ganesh (1995) was used and is described in detail in the statistical appendix.

- **Step 1** – Choice of potential confounding (prognostic) factors (X_1, \dots, X_4), and strata (j) for each factor. Cases were allocated to 64 strata within four factors: (i) age (four levels: <40, 40–49, 50–59, 60+), (ii) stage of disease (four levels: I, II, III and IV, unknown), (iii) cancer-directed treatment (two levels: yes, no) and (iv) place of residence (two levels: Muang and surrounding districts, other). Muang district is in the centre of Khon Kaen province where Khon Kaen city is located.
- **Step 2** – Classification of study subjects into two main categories: those with complete follow-up and those lost to follow-up.

At a given survival time (annual), ' t ' = 1–5 (say), the subjects in each stratum (n_{ij}) were classified into two groups: (1) those 'completely followed up', denoted by n_{ij}' , comprising those dead (d_{ij}) during the interval (i) or alive (w_{ij}) at the end of the annual interval, and (2) those 'lost to follow-up', denoted by l_i , who were last known to be alive in the annual interval and status unknown thereafter.

- **Step 3** – Computation of probability of death (q_{ij}') for all i and j for factors X_1, \dots, X_4 among cases with complete follow-up ($n_{ij}' = n_{ij} - l_i$). The probability of death (q_{ij}') at each annual interval i , was estimated by means of a logistic regression model, using cases with complete follow-up only (n_{ij}'), with all factors (j) taken into account simultaneously in the model. Stata version 7.0 (2001) software was used to estimate the regression coefficients.
- **Step 4** – Computation of expected deaths (d_{ij}') among cases lost to follow-up (l_i). The expected deaths (d_{ij}') among the group of cases lost to follow-up (l_i) were estimated by assigning the same probability (q_{ij}') of death.
- **Step 5** – Computing the loss-adjusted survival for each interval i . The computation of the conditional probability of dying (q_i), conditional probability of surviving (p_i) and the cumulative probability (P_i) of surviving the current and subsequent annual intervals of time are estimated by accumulating the numbers d_{ij} , d_{ij}' , n_{ij} , n_{ij}' over the confounders, j , and proceeding under the modified actuarial framework of generating life table, as described in the statistical appendix.

RESULTS

A total of 601 (95.4%) out of 630 cases of cancer of the cervix diagnosed during 1985–1990 were included in the survival study; all patients were followed to the end of 1995 or later. In all, 83% were diagnosed microscopically. Table 1 shows the distribution of cases by age, stage, treatment received and place of residence. In

Table 1 Number of cases, proportion and risk (odds ratio, OR) of death and loss to follow-up at 5 years from the index date and 95% confidence interval (CI) by factors studied

Factors studied	Number of cases	Proportion at 5 years from index date		Odds ratio (OR) and 95% CI	
		Lost (%)	Dead (%)	Lost OR ^a	Dead ^b OR ^a
All cases	601	27.6	36.4		
(1) Age group					
<40	122	27.9	24.6	1.0	1.0
40–49	194	24.2	35.1	0.9 (0.5–1.5)	1.5 (0.8–2.7)
50–59	158	29.1	36.7	1.2 (0.7–2.0)	2.0 (1.1–3.7)
60+	127	29.9	49.6	1.2 (0.7–2.1)	3.5 (1.8–6.9)
(2) Stage of diseases					
Stage I	93	23.7	20.4	1.0	1.0
Stage II	134	28.4	29.1	1.2 (0.7–2.2)	1.8 (0.9–3.6)
Stage III and IV	222	21.6	53.6	0.8 (0.4–1.5)	5.0 (2.7–9.5)
Stage unknown	152	37.5	27.6	1.4 (0.7–2.6)	1.5 (0.7–3.1)
(3) Treatment					
Received treatment	428	22.4	36.2	1.0	1.0
No treatment	173	39.9	37.0	2.0 (1.3–3.1)	2.0 (1.2–3.4)
(4) Residency					
Muang and surrounding					
Districts	274	28.1	32.9	1.0	1.0
Other districts	327	26.9	39.4	0.8 (0.6–1.2)	1.5 (1.0–2.3)

^aORs of each factor adjusted for all other factors in the table. ^bEstimated among those with complete follow-up only.

all, 316 cases (53%) were aged under 50 years, and 127 (21%) were over 60; 93 cases (15.5%) were stage I, 22.3% stage II, 31.5% stage III, 5.5% stage IV and 25.3% were of unknown stage at diagnosis. A total of 45.6% of cases were residents of Muang and surrounding districts; 71.2% of patients received treatment through either surgery or radiation or chemotherapy.

Risk of loss to follow-up and death

The proportion and risk (odds ratio) of death and loss to follow-up at 5 years from the index date, by prognostic factors, are presented in Table 1. The proportion of patients lost to follow-up during the 5-year period was 27.6%, and of dying was 36.4%. The risk of loss to follow-up varied 1.3-fold by age at diagnosis, 1.6-fold by stage of disease and 1.2-fold by place of residence; the risk of loss to follow-up among cases not treated was two-fold higher than those treated.

The risk of death increased 3.5-fold with increasing age at diagnosis, and five-fold with stage of disease ($P < 0.001$), with the highest risk observed in stages III and IV. Those with stage unknown also had a higher risk of death than those in stage I (OR = 1.5). Patients with no treatment had a two-fold higher risk of death and patients who lived far away from the centre of the province had a 50% higher risk than those who lived nearby.

Survival from cervix cancer (actuarial and loss-adjusted)

The observed (actuarial) survival at 5 years was 56.8% (Table 2). During this period, 27.6% of cases were lost to follow-up; 13.3% in the first year, 5.1% of those remaining in the second and third years, and 19.3% of the remainder in the fourth and fifth years (Table 2).

Adjustment for loss of follow-up gave an estimated survival of 54.7% at 5 years from index date, 2.1% units less than the observed (actuarial) survival. This suggests that the patients who were lost to follow-up had a higher mortality than assumed in the actuarial method of survival analysis, in which such deaths occur at the

same rate as among those with complete follow-up. Table 2 also gives the estimate of loss-adjusted survival by age group, stage, treatment and residence, each adjusted for differential loss to follow-up by the other three factors.

Age An inverse relationship between survival and age at diagnosis was evident: Patients aged less than 40 years had the best survival and patients aged more than 60 years had the poorest survival by both estimation methods. The degree of bias introduced into the actuarial estimate by differential loss to follow-up was small, in the range of 1.2–2.9% units, and the variation by age was somewhat less than indicated by the actuarial estimates.

Stage of disease Patients with stage I had the best survival (74.6%) and stage III and IV had the poorest survival (38.2%). The reduction in the differences of survival between loss-adjusted and actuarial estimates was the highest in patients with unknown stage (3.5% units) and the smallest in patients with stage III and IV disease (0.8% units).

Treatment Patients who received treatment had better survival (57.5%) than those who had not (47.5%). The reduction in the differences of survival between loss-adjusted survival and actuarial estimates was higher in the untreated (2.4% units) than in those who received treatment (1.7% units).

Place of residence Patients who lived in Muang and surrounding districts had better survival (58.6%) than patients who lived in other districts (51.4%). Loss-adjusted survival revealed a reduction in estimated survival compared with that estimated by the actuarial method, for both residence groups. The difference in survival estimates was 2.5% units for residents of Muang and surrounding districts and 1.8% units for those living in other districts.

These small changes in the estimate of survival following the loss-adjustment procedure indicate the presence of a small bias in

Table 2 Number of cases, proportion lost to follow-up at varying intervals of time and 5-year cumulative absolute and loss-adjusted survival of factors studied

Factors studied	No. of cases	% lost to follow-up among persons at risk of death at varying lengths of time (i) from index date			% Absolute survival	
		< 1 year	1 ≤ i < 3 years	3 ≤ i < 5 years	Actuarial	Loss-adjusted ^a
All cases	601	13.3	5.1	19.3	56.8	54.7
(1) Age group						
< 40	122	13.1	4.1	17.9	71.0	68.1
40–49	194	11.3	4.8	16.4	59.7	57.8
50–59	158	13.9	6.6	19.8	55.7	53.4
60+	127	15.8	4.8	26.9	38.8	37.6
(2) Stage of diseases						
Stage I	93	7.5	1.2	19.4	77.1	74.6
Stage II	134	9.7	9.0	18.3	65.5	63.0
Stage III and IV	222	9.0	5.8	20.2	39.0	38.2
Stage unknown	152	26.3	3.2	19.2	63.3	59.8
(3) Treatment						
Received treatment	428	6.8	6.4	17.2	59.2	57.5
No treatment	173	29.5	0.0	27.7	49.9	47.5
(4) Residency						
Muang and surrounding districts	274	13.1	5.2	19.6	61.1	58.6
Other districts	327	13.5	5.1	19.0	53.2	51.4

^aEach factor adjusted for differential loss to follow-up by other factors in the table.

the actuarial estimate, resulting from the higher mortality among cases lost to follow-up, than under the actuarial assumption.

DISCUSSION

The fundamental step in carrying out an end result study is to ensure good and complete follow-up of patients. The actuarial (life table) method uses information from all subjects, including those censored before 5 years follow-up or death. Losses to follow-up and withdrawals may have different effects on the estimates of survival. The actuarial survival rate gives an unbiased estimate of true survival only if censorship has the same distribution between the groups being compared (Hakulinen, 1982) and is independent of risk of the outcome studied (Ganesh, 1995). The bias in the estimation of survival probability is dependent on both the magnitude and nature of losses to follow-up, and may be in either direction. For example, the true probability of death of patients lost to follow-up may be greater than assumed if patients with poor prognosis are more likely to be lost. In these circumstances, the actuarial survival estimate is biased and too high.

If the vital status of all the cases included in a survival study is known at the closing date, the estimation of survival probability by the actuarial method is straightforward and unbiased. In the present study, all subjects could be potentially followed for at least 5 years, so that there were no withdrawals, and all censoring was due to loss to follow-up. In developing countries, it is difficult to obtain complete follow-up information for all patients for various reasons. Typically, cancer patients no longer being followed up in hospital must be traced by active methods, involving postal enquires or home visits. Patients frequently migrate from their usual place of residence to that of their relatives and the hospital/medical centre may not be informed of the change in address. This makes tracing of patients at home difficult, since the new contact address must be obtained from other sources, neighbours or friends, for example. Migration is typically related to the recurrence of the disease, that is, with factors of prognostic significance; its magnitude depends on the nonrandom nature and the extent of the loss to follow-up. It is therefore important in any survival study to ascertain not only the extent of loss to follow-up, but also its independence of the probability of death.

REFERENCES

- Breslow N, Day NE (1980) *The Analysis of Case-Control Studies. Statistical Methods in Cancer Research*. Vol 1. IARC Scientific Publications No 32: Lyon, France
- Cutler SJ, Ederer F (1958) Maximum utilization of the life table method in analyzing survival. *J Chronic Dis* 8: 699–712
- Dickman P, Hakulinen T, Voutilainen E (2002) *Surv3 Relative Survival Analysis Program Version 3.00b2*. Finnish Cancer Registry: Helsinki, Finland
- Ederer F, Lillian MA, Cutler SJ (1961) The relative survival rate: A statistical methodology. *Natl Cancer Inst Monogr*. 6: 101–121
- Ganesh B (1995) *Effect of Loss to Follow-up in Estimating Survival Rates*. Ser A, Vol. 440. Acta Universitatis Tamperensis: Tampere
- Hakulinen T (1982) Cancer survival corrected for heterogeneity in patient withdrawal. *Biometrics* 38: 933–942

STATISTICAL APPENDIX

Loss-adjusted survival

The procedure for estimating loss-adjusted survival by the stratified method can be described step by step in the actuarial survival estimation framework. However, there is also a different approach which integrates estimation by a regression technique and the life table approach. It is the latter that is sequentially described here:

The first step in deciding whether bias in the actuarial estimate of survival is likely is to examine whether loss to follow-up varies according to prognostic variables such as age, stage, residence and treatment group. Computation of loss-adjusted survival (Ganesh, 1995) then takes into consideration such differential losses, by assuming that patients lost to follow-up within strata defined by these variables have the same probability of death as those still remaining under observation and belonging to the same stratum. It is reasonable to expect survival experience in patients lost to follow-up and with complete follow-up to be more similar within a prognostic group, than when all patients are considered together. The difference between the crude actuarial survival and the loss-adjusted value indicates the magnitude of the effect of differential loss to follow-up.

The small difference between the absolute (actuarial) survival and the loss-adjusted survival observed in this study is much less than in other studies (Ganesh, 1995; Mathew, 1996). Large differences in LAR and actuarial estimates have been found in hospital-based series of patients, coming from a wide geographic area, where follow-up of patients no longer attending hospital clinics by house visits was impractical and no postal enquiries were made. In contrast, an international comparison of actuarial and loss-adjusted survival of cervix cancer cases from different population-based cancer registries in developing countries (Swaminathan *et al.* 2002) found that the maximum difference was 4.1%, with a loss to follow-up of 44% and presence of nonrandomness. The observation was not confined to cancer of the cervix; differences for other sites like female breast (data from six registries from developing countries) and larynx (data from Chennai and Mumbai cancer registries) were of similar (small) size. This may be mainly because of the integration of mortality data collection into the case-finding operations of population-based cancer registries (on an annual basis in Khon Kaen). It confirms the finding of the present study, that in a population-based series the assumption of independency of loss to follow-up and death was reasonable, so that calculation of survival by the actuarial method without adjusting for losses to follow-up is likely to have resulted in no material bias in the estimates. However, this is not true in general; the experience of hospital-based series in particular indicates that bias may be considerable, and requires appropriate adjustment of survival estimates.

- Kogevinas M, Porta M (1997) Socioeconomic differences in cancer survival: a review of the evidence. In *Social Inequalities and Cancer*, Kogevinas M, Pearce N, Susser M and Boffetta P (eds) pp. 177–206, IARC Scientific Publication No 138: Lyon, France
- Mathew A (1996) *Removing Bias in Cancer Survival Estimates by Active Follow-up and Information on Determinants of Loss to Follow-up*. Ser A, Vol 525. Acta Universitatis Tamperensis: Tampere
- STATA (2001) *Stata Software 2001*. College Park, TX
- Swaminathan R, Sankaranarayanan R, Hakama M, Shanta V (2002) Effect of loss to follow-up on population based cancer survival rates in developing countries. *Int J Cancer* 100(Suppl. 13): 172 (18th UICC Cancer Congress, 30 June–5 July 2002, Oslo, Norway – Abstract book)

In the follow-up interval i in prognostic stratum j , there will be n_{ij} patients alive at the beginning of the interval, of whom d_{ij} will die, w_{ij} will be withdrawn alive because of the closing date of follow-up and l_{ij} will be lost to follow-up during the interval. Assuming for simplicity that the potential follow-up exceeds i intervals, $w_{ij} = 0$. The number with complete follow-up, n_{ij}' , is

given by

$$n_{ij}' = n_{ij} - l_{ij}$$

To overcome the problems in estimation by multifactorial methods caused by sparse numbers in the stratified approach, loss-adjusted survival can also be estimated by logistic regression methods (Breslow and Day, 1980).

The proportion dying with complete follow-up, q_i' , given the prognostic factors x_i, \dots, x_k , is first estimated for patients not lost to follow-up, n_{ij}' , in the interval i :

$$q_i' = \frac{\exp(\mu_i')}{1 + \exp(\mu_i')}$$

where $\mu_i' = \beta_{0i} + \beta_{1i}x_{1i} + \dots + \beta_{ki}x_{ki}$, the hazard in interval i given the prognostic factors. The proportion of deaths can be estimated for each level of any prognostic variable x_i adjusting for the effect of the other prognostic variables. This is done for every interval i .

The expected number of deaths among patients lost to follow-up is then computed as $d'_{ij} = q'_{ij}l_{ij}$, and the expected proportion of deaths among all n_{ij} cases is

$$q_{ij} = \frac{d_{ij} + d'_{ij}}{n_{ij}} = \frac{D_{ij}}{n_{ij}}$$

The procedure is repeated for the next interval ($i = i + 1$) with $n_{(i+1)j}' = n_{ij} - d_{ij} - l_{(i+1)j}$ and with $l'_{(i+1)j} = l_{(i+1)j} + l_{ij} - d'_{ij}$, and with $d'_{(i+1)j} = q'_{(i+1)j}l'_{(i+1)j}$ and for the other prognostic strata.

Accumulating over prognostic strata will result in an annual loss-adjusted rate:

$$q_i(\text{Loss Adjusted}) = \frac{\sum_j D_{ij}}{\sum_j n_{ij}}$$

and the cumulative loss-adjusted (crude) survival rate is

$$p_i(\text{Loss Adjusted}) = (1 - q_1)(1 - q_2) \dots (1 - q_i)$$

If there are withdrawals (w_{ij}), the normal actuarial approximation is

$$q_{ij} = \frac{D_{ij}}{n_{ij} - \frac{1}{2}w_{ij}}$$

The approach corresponds to adjustment by stratification.

The process described for the stratified analysis will be applied for d'_{ij} , $q_i(\text{LAR})$ and ultimately for $P_i(\text{LAR})$, by repeating the estimation of $q_i(\text{LAR})$ over the first i intervals.

Chapter 3

Loss-adjusted hospital and population-based survival of cancer patients

Ganesh B, Swaminathan R, Mathew A, Sankaranarayanan R and Hakama M

Abstract

This chapter presents formulae that methodologically adjust for losses, and gives examples describing magnitude of bias in survival estimates without such adjustment. Loss-adjusted survival is estimated under the assumption that survival of patients lost to follow-up is the same as that for patients with known follow-up time and similar characteristics of different prognostic factors at first entry. The observed number of losses to follow-up is then relocated into expected numbers of death and survivors on this basis. Standard methods, such as the actuarial one, are then applied with the sum of observed and expected outcome events. A total of 336 hospital series of treated new breast cancer cases from Mumbai with 24% lost to follow-up revealed a substantial bias of 7 per cent units for 3-year survival estimated with (54%) and without (61%) loss-adjustment. Stepwise adjustment of losses established that increasing the number of prognostic factors explained the bias better. Population-based series comprising 13 371 cases of top ranking cancers from Chennai, with loss to follow-up ranging from 7–24%, revealed negligible bias, ranging from 0–2% in 5-year survival by the loss-adjusted approach for different cancers. Data source seems to affect the need for loss-adjustment, and the loss-adjusted approach is recommended when hospital-based cancer registry data of a low- or medium-resource country are used to evaluate the outcome of cancer patients.

Introduction

Cancer survival is the main indicator of outcome of cancer health services or treatment, and an important component in maintaining cancer control activities [1]. Cancer registries have long served as potential sources of data for estimating survival. Hospital-based cancer registries usually report survival of a selected series of treated patients that are registered in a hospital or group of hospitals without specific coverage of geographical area or background population. On the other hand, population-based cancer registries, which include all incident cases treated or not from a specific geographical area, usually report average survival in specific regions. Cancer survival reported from both settings may have different perspectives, but estimation of survival rates is routinely done using standard life table approaches such as the actuarial [2] or Kaplan-Meier [3] methods.

The actuarial method [2] of estimating survival by follow-up time allows utilization of all information independent of the length of follow-up of an individual patient, so that even recently diagnosed

patients contribute to long-term survival. Patients who have a potential follow-up shorter than the time of the maximum estimated survival are "censored" cases. Censored cases are usually withdrawals, surviving at date of last follow-up: this date can be either individual for each patient or a common closing date for all patients. However, censorship in terms of losses to follow-up takes place if follow-up fails before this potential withdrawal. There is a qualitative difference between these two groups of censored cases.

Losses to follow-up may cause major bias. This holds true if the losses are common and correlated with the patient prognosis or survival. In most low- or medium-resource countries, such losses are common due to deficiencies in health infrastructure and recording of health statistics. The losses are also likely to be related to the patient's prognosis: low social status is related to lack of continuous patient surveillance; extent of disease is related to the motivation of follow-up, etc. Hence, this correlation, explained by information on prognostic factors, can be utilized to correct survival estimates.

A method to estimate loss-adjusted survival rates corrected, for possible bias due to losses to follow-up, is described here through two examples, one each from hospital-based and population-based registry settings. The loss-adjusted survival results are compared with the crude actuarial estimate to demonstrate the magnitude of bias.

Methods

Follow-up

Follow-up was carried out by passive and active methods. The passive approach was by data linkage either with patients' records on regular follow-up at the outpatient clinic and/or with mortality data from the vital statistics division. The active approach was by contacting the patients or their families directly by means of postal/telephone/e-mail/house visit enquiries for information on survival status.

Determinants of loss to follow-up or survival

Categorical factors (like age, sex, literacy status, tumour stage, treatment, etc.), each with reference and subcategory levels, that have the potential to influence either follow-up (complete or lost to follow-up) or survival (alive or dead) were first determined by using test of proportions (univariate only), logistic regression (unifactorial or multifactorial) or Cox proportional-hazard model (univariate or multifactorial using survival time information). A differential pattern of follow-up or survival outcome, either between factors or within subcategories of factors, would indicate an association of non-random nature.

Estimation of loss-adjusted survival rate - stratified method

The life table method estimates annual survival during a given follow-up year by specifying four types of events including the outcome experienced by the patient: surviving throughout the year; dying (outcome) during the year; withdrawn alive, where patient was known to be alive at closing date of follow-up; and loss to follow-up, where the known survival time terminates during the follow-up year, but before closing date. Unlike traditional survival analysis, which grouped withdrawals and losses together, the proposed method for estimating loss-adjusted survival differentiated the two. For the time being, methods are developed for potential follow-up time of all subjects equalling the time for which survival is estimated. In other words, potential follow-up time for all cases would have to be five years to estimate 5-year loss-adjusted survival rate.

Every prognostic stratum is composed of a unique combination of subcategories of all identified determinants of follow-up or survival. In the estimation of loss-adjusted survival, it is assumed that those lost to follow-up in specific prognostic stratum have the same probability of death as others still remaining under observation and belonging to the same stratum. At any given follow-up time, the observed numbers of losses to follow-up in each stratum are relocated into expected numbers of deaths, withdrawals and survivors on the basis of observed survival in those without loss to follow-up in the same stratum. The actuarial method, or any other, is then applied to the sum of observed and expected events.

In the follow-up interval i in prognostic stratum j , there will be n_{ij} patients alive at beginning of interval, of whom d_{ij} will die, w_{ij} will be withdrawn alive and l_{ij} will be lost to follow-up during the interval. Since potential follow-up exceeds i intervals for all patients, $w_{ij} = 0$. The number with complete follow-up, n'_{ij} , is then given by:

$$n'_{ij} = n_{ij} - l_{ij}.$$

The proportion dying with complete follow-up, q'_{ij} given the prognostic factors x_j, \dots, x_k , is first estimated for patients not lost to follow-up, n'_{ij} , in the interval i :

$$q'_{ij} = \frac{d_{ij}}{n'_{ij}}.$$

The expected number of deaths in patients lost for follow-up in interval i is:

$$d'_{ij} = q'_{ij} l_{ij}$$

and the expected proportion of deaths in the n_{ij} cases is:

$$q_{ij} = \frac{(d_{ij} + d'_{ij})}{n_{ij}} \\ = \frac{D_{ij}}{n_{ij}}.$$

The procedure is repeated for the next interval ($i = i + 1$) as follows:

$$n'_{(i+1)j} = n_{ij} - D_{ij} - l_{(i+1)j} \quad \text{and with}$$

$$l'_{(i+1)j} = l_{(i+1)j} + l_{ij} - d'_{(i+1)j} \quad \text{with}$$

$$d'_{(i+1)j} = q'_{(i+1)j} l'_{(i+1)j} \quad \text{and for the other prognostic strata.}$$

Accumulating over prognostic strata will result in an annual loss-adjusted rate:

$$q_i(\text{Loss Adjusted}) = \frac{\sum_j D_{ij}}{\sum_j n_{ij}}$$

and the cumulative loss-adjusted survival probability is:

$$P_i(\text{Loss Adjusted}) = (1 - q_1)(1 - q_2) \dots (1 - q_i).$$

Logistic regression approach to estimate expected deaths among loss to follow-up

The correction of bias in survival estimation adjusted for loss to follow-up is optimal when it is determined by including as many factors as possible. An increase in number of determinants (factors with subcategories) of follow-up or survival would result in a corresponding increase in the number of prognostic strata. Cross-tabulation of all of these factors simultaneously would require adequate sample size to keep a majority of prognostic strata non-empty. Adjusting all factors simultaneously by logistic regression is a simplification of the computational procedure to estimate expected deaths among lost to follow-up and offers maximal effect in reducing the bias.

The proportion dying in the $\sum_j n'_{ij}$ patients followed completely during the interval is:

$$q'_i = \frac{\exp(\mu'_i)}{(1 + \exp(\mu'_i))}$$

where

$$\mu'_i = \beta_{0i} + \beta_{1i}x_{1i} + \beta_{2i}x_{2i} + \dots + \beta_{ki}x_{ki}$$

is a linear combination of the determinant or prognostic factors. The above methods are described in detail elsewhere [4,5].

Other approaches

Loss-adjusted survival can also be estimated using the Kaplan-Meier approach [6]. Stratum-specific expected deaths are estimated and the Kaplan-Meier curve is corrected at time points when the expected deaths occur.

Results

Example 1: Hospital-based cancer registry series

A total of 336 new cases of female breast cancer cases that were diagnosed and received complete treatment at Tata Memorial Hospital, Mumbai (Bombay), India, in 1985 and followed-up until 1988 formed the study population. These cases were allocated to 64 strata involving four factors associated with follow-up or prognosis: age (in completed years: <45, 45–54, 55–64, 65+ years); stage of disease (TNM staging classification: I, II, III, IV); type of treatment (chemotherapy: without, with); place of residence (Mumbai: residents, non-residents). Outcome event with respect to follow-up was loss to follow-up <3 years from diagnosis, and outcome event for loss-adjusted survival was death due to any cause.

Patients below 55 years of age comprised 65%, with an overall mean of 49 years (Table 1). There was an equal distribution of resident and non-resident patients from Mumbai city. A majority were diagnosed in stage II (48%) followed by stage III (37%) of the disease. About 58% of the patients were treated with either surgery or radiotherapy or in combination but not with chemotherapy, while the remaining 42% were treated with chemotherapy either alone or in combination with other modalities. Differential pattern of proportion (%) or risk (odds ratio) of loss to follow-up by different prognostic factor categories was forthcoming. The proportion of patients lost to follow-up was not very different between subcategories of age and type of treatment, with 0 to 30% increased risk over corresponding reference categories that was statistically not significant. The proportion lost to follow-up was doubled among non-residents versus residents of Mumbai, with two- to three-fold increased risk that was statistically significant. The risk was two to three times higher among stage III or IV patients and 50% higher among stage II compared to stage I patients, but not statistically significant (Table 1). The findings suggest an association between these prognostic factors and loss to follow-up.

The data was further analysed to estimate loss-adjusted survival by stratification of two or three factors at a time and by logistic regression approaches. Survival was estimated at the end of 3-year follow-up by actuarial method without and with adjustment for loss to follow-up (Table 2). The 3-year survival obtained by loss-adjustment showed lower survival compared to rates obtained by standard actuarial assumption without specific adjustment for loss to follow-up. The bias in survival estimation is represented as the difference in per cent units of survival rates (%) without and with loss-

adjustment for each factor. This varied from 5.4 for patients aged 55 to 64 years to 8.6 for those aged <45 years. The bias was lesser among Mumbai residents (3.2) than non-residents (8.8). Three-year loss-adjusted survival was higher among residents (56.2%) than non-residents (54.4%), but this was the opposite for corresponding survival figures without loss-adjustment (59.4% and 63.2%), respectively. A decrease in survival (Table 2) and increase in proportion of lost to follow-up (Table 1) with severity of disease was forthcoming, which indicated a positive association between risk of dying and loss to follow-up in all disease stages. Loss-adjusted survival was greater in stage I patients, but lesser in other stages, compared to respective survival estimates without loss-adjustment. Following the elimination of bias by loss-adjustment, the difference in loss-adjusted survival between stages I and III patients increased from 51 per cent units to 61 per cent units (Table 2). The proportion of deaths in the chemotherapy group was twofold more than in the non-chemotherapy group. The comparison between actuarial and loss-adjusted survival showed that the adjusted unbiased difference between the two groups was bigger (43 per cent units) than the unadjusted ones (38 per cent units).

The variable extent of bias in survival estimation that could be elicited in the presence of loss to follow-up by utilizing information from one to four prognostic factors is shown stepwise for all cases in Table 3. The unadjusted actuarial 3-year survival was 61%. The loss-adjustment yielded a decrease of 7 per cent units in survival when all four prognostic factors were considered simultaneously by logistic regression method. The stepwise introduction of each of the prognostic factors into the adjustment procedure, by stratified method of estimating loss-adjusted survival, increased the correction of bias as follows: 1.7 per cent units when adjusted only for residential status; 2 per cent units when age was added; 3.8 per cent units when stage was added to the previous two factors; and 4.7 per cent units when all factors were adjusted.

Example 2: Case series from Chennai population-based cancer registry

A total of 13 371 cases comprising cancers of the uterine cervix (3134), female breast (1923), stomach (1845), oesophagus (1403), lung (1237), mouth (1202), lymphomas (768), tongue (670), leukaemias (668), and of ovary (521) ranked within the top ten in

Table 1. Number and proportion (%) of patients and losses at 3 years and risk (odds ratio) of loss to follow-up with 95% confidence interval by patient characteristics among female breast cancer patients diagnosed in Tata Memorial Hospital, Mumbai, India, in 1985 and followed through 1988

Patient characteristics	Patients (n=336)		Lost to follow-up (n=80; 24%)		Odds ratio (95% CI)
	Number	% ^a	Number	% ^b	
Age at diagnosis					
≤ 44 years	101	30	22	22	1.0*
45–54	117	35	29	25	1.2 (0.6–2.3)
55–64	77	23	19	25	1.2 (0.6–2.5)
65+ years	41	12	10	24	(0.5–2.9)
Residential status (Mumbai city)					
Residents	169	50	26	15	1.0*
Non-residents	167	50	54	32	2.6 (1.5–4.6) [§]
Stage of disease (TNM summary)					
I	29	9	4	14	1.0*
II	160	48	30	19	1.5 (0.5–5.6)
III	126	37	40	32	2.9 (0.9–10.6)
IV	21	6	6	29	2.5 (0.5–12.9)
Treatment					
With chemotherapy	194	58	42	22	1.0*
Without chemotherapy	142	42	38	27	1.3 (0.8–2.3)

^a Percentage of total breast cancer cases;

^b Percentage of total cases in respective categories;

CI: Confidence interval;

* Reference category;

[§]p=0.05.

Table 2. Number and proportion (%) of patients and deaths and comparison of 3-year survival with and without adjustment for loss to follow-up by patient characteristics among female breast cancer patients diagnosed in Tata Memorial Hospital, Mumbai, India, in 1985 and followed through 1988

Patient characteristics	Number of patients	Deaths		3-year survival %	
		Number	% ^a	Actuarial assumption	Loss-adjusted by logistic regression*
Age at diagnosis					
≤ 44 years	101	34	34	60.1	51.5
45–54	117	42	36	56.7	48.7
55–64	77	20	26	67.7	62.3
65+ years	41	12	29	65.4	58.5
Residential status (Mumbai city)					
Residents	169	60	35	59.4	56.2
Non-residents	167	48	29	63.2	54.4
Stage of disease (TNM summary)					
I	29	2	9	92.2	93.2
II	160	36	22	74.4	71.2
III	126	55	44	41.2	31.8
IV	21	15	71	0.0	0.0
Treatment					
With chemotherapy	194	39	20	76.6	71.2
Without chemotherapy	142	69	48	38.1	28.2

^a Percentage of total cases in respective categories;

* Adjusted for other factors in the table.

Table 3. Comparison of 3-year survival without loss-adjustment by actuarial assumption, stepwise loss-adjustment of factors using stratified method and loss-adjustment using all factors together by logistic regression for all female breast cancer patients diagnosed in Tata Memorial Hospital, Mumbai, India, in 1985 and followed through 1988

Loss-adjustment of factors	3-year survival %
Without loss-adjustment and using actuarial assumption only	61.2
Loss-adjustment done by stratification	
Residential status	59.5
Residential status and age at diagnosis	59.2
Residential status, age at diagnosis and stage of disease	57.4
Residential status, age at diagnosis, stage of disease and treatment	56.5
Loss-adjustment done by logistic regression	
Residential status, age at diagnosis, stage of disease and treatment	54.5

the Population-Based Cancer Registry, Chennai, India, during 1990–1996 and followed-up until 2001 formed the study population.

The determinants of loss to follow-up at less than 5 years from diagnosis for each site were identified using Cox proportional-hazard model by following the method outlined in Chapter 2 of this publication. Five-year loss-adjusted absolute survival of patients

through stratified method was estimated by allocating cases to 128 strata defined by 4 factors (with reference and subcategories): age at diagnosis (<45, 45–54, 55–64 and 65+ years); literacy status based on years of education (Nil, ≤ 5, 6–12 and ≥ 12 years); clinical extent of disease as a surrogate for tumour stage (localized, regional, distant metastasis and unknown); treatment status (no or unknown and yes). Outcome event was death due to any cause.

Table 4 gives the proportion of cases lost to follow-up and comparison of 5-year absolute survival estimated with and without adjustment for loss to follow-up for each cancer site. The losses ranged between 7% (oesophagus) and 24% (ovary) for different sites. Loss-adjusted survival was consistently lesser than the corresponding unadjusted estimate for all sites. Bias in survival estimation in the presence of non-random loss to follow-up, expressed in terms of absolute difference between survival (%) estimates obtained with and without loss-adjustment was minimal, ranging between 0.2 to 1.7 per cent units for different cancer sites.

Discussion

The success of cancer treatment is, as a rule, measured by survival. Population-based survival reflects the availability, development of and accessibility to cancer health services in a region. Survival based on hospital series reflects the impact of clinical services specific to the hospital. In both instances, high-level completeness of ascertainment of mortality data is an important prerequisite, and when such completeness cannot be assured, survival rates should be carefully interpreted [7,8].

Conventionally, estimation of survival was done using life table approaches by either actuarial [2] or Kaplan-Meier [3] methods. Both methods utilize observed survival time independently of whether it ends at the death of a patient. Patients withdrawn alive at closing date provide censored information that is unbiased, since closing date is independent from probability of death. If this is not true,

Hakulinen [9] and Brenner [10] give means to adjust for withdrawal pattern and to correct for effects of improvement of survival by time.

Losses to follow-up because of reasons other than closing date (e.g., migration) are often few in developed countries and are dealt with identically as withdrawals. This is not justified if the losses are many and are correlated with risk of death. Distance from clinical care facility increases the likelihood of not undergoing a follow-up examination, as does serious morbidity and poverty. The factors in failure to obtain follow-up data are the same. Therefore, it is likely that patients lost to follow-up have poor prognosis and could not be compared with those under follow-up and surveillance. The direction in bias may also be the other way: those lost to follow-up have a better survival than those under follow-up, as was shown in our example on stage I breast cancer hospital series patients.

Our example from a hospital series shows that the bias due to losses may be substantial. Mathew[6] showed similar differences by applying loss-adjustment in the Kaplan-Meier survival method for hospital series ovarian cancer patients. Much of the original deficiencies in the hospital data were, however, removed by active follow-up using a postcard enquiring the vital status of patient. Only marginal adjustment effect appeared after the enquiry. However, in the example involving breast cancer hospital series, a large bias still existed after such attempts of active follow-up. On the other hand, the example involving population-based series of several cancers revealed negligible bias. In both

Table 4. Number of incident cases, proportion (%) lost to follow-up and comparison of 5-year absolute survival with and without loss-adjustment for top-ranking cancers in a population-based cancer registry, Chennai, during 1990–1996 and followed through 2001

Cancer site/type	Number of incident cases	Lost to follow-up %	5-year survival %		Absolute difference in survival
			No loss-adjustment	With loss-adjustment	
Cervix	3134	21.8	52.1	50.4	1.7
Breast	1923	20.7	39.5	39.1	0.4
Stomach	1845	8.0	9.4	8.7	0.7
Oesophagus	1403	6.7	7.7	7.5	0.2
Lung	1237	7.8	8.2	8.1	0.1
Mouth	1202	11.6	30.1	29.1	1.0
Lymphomas	768	11.5	26.5	25.6	0.9
Tongue	670	13.0	20.2	18.9	1.3
Leukaemias	668	8.2	19.8	19.2	0.6
Ovary	521	24.0	25.7	24.2	1.5

instances, loss-adjusted survival was lesser than the actuarial estimate without adjustment indicating that under-ascertainment of deaths among loss to follow-up cases may be the problem. Most population-based cancer registries are based on systems that integrate linkage or collection of mortality data as a routine and hence result in small differential bias only [5]. Hence, the data source seems to affect the need for loss-adjustment, and the problem may be more substantial in hospital-based cancer registries and clinical series. The loss-adjusted approach is likely to be useful especially when hospital-based cancer registry data of a low- or medium-resource country are used to evaluate the outcomes of cancer patients.

One may conclude that if routine follow-up is poor, the first priority is to increase the actual follow-up visits on humanitarian and scientific grounds. The second is to improve the data by instituting rigorous active follow-up measures. The improvement of data by these means may indirectly improve routine follow-up activity. Analytical methods to correct the survival data with adjusting for losses are to be used in surveillance and evaluation and in scientific comparisons. However, such means do not directly improve human health, but have the potential to improve the organization itself.

References

1. WHO/IARC. *Cancer Statistics: Report of a WHO/IARC Expert Committee. WHO Technical Report Series No. 632.* World Health Organization, Geneva, 1979.
2. Cutler SJ, Ederer F. Maximum utilization of the life table method in analysing survival. *J Chron Dis* 1958;8: 699–712.
3. Kaplan EL and Meier P. Non parametric estimation from incomplete observation. *J Amer Stat Assoc* 1958;53:457–481.
4. Ganesh B. *Effect of loss to follow-up in estimating survival rates. Acta Universitatis Tamperensis, Ser A, Vol 440.* University of Tampere, Tampere, 1995.
5. Sriamporn S, Swaminathan R, Parkin DM, Kamsa-Ard S, Hakama M. Loss-adjusted survival of cervix cancer in Khon Kaen, Northeast Thailand. *B J Cancer* 2004;91(1):106–110.
6. Mathew A. Removing bias in cancer survival estimates by active follow-up and information on determinants of loss to follow-up. *Acta Universitatis Tamperensis, Ser A, vol 525.* University of Tampere, Tampere, 1996.
7. Brenner H, Hakulinen T. Implications of incomplete registration of deaths on long-term survival estimates from population-based cancer registries. *Int J Cancer* 2009;125:432–437.
8. Swaminathan R, Rama R, Shanta V. Lack of active follow-up of cancer patients in Chennai, India: implications for population-based survival estimates. *Bull World Health Organ.* 2008;86:509–515.
9. Hakulinen T. Cancer survival corrected for heterogeneity in patient withdrawal. *Biometrics.* 1982;38: 933–942.
10. Brenner H, Gefeller O. An alternative approach to monitoring cancer patient survival. *Cancer.* 1996;78: 2004–2010.

COMMENTARY

Global Variations in Cancer Survival

R. Sankaranarayanan, M.D.¹

R. Swaminathan, M.S.C.²

R. J. Black, M.A.¹

for the Study Group on Cancer Survival in Developing Countries

¹ Unit of Descriptive Epidemiology, International Agency for Research on Cancer, Lyon, France.

² Cancer Institute (WIA), Adyar, Madras, India.

Population-based cancer registries from Algeria, China, Costa Rica, Cuba, India, the Philippines, and Thailand are collaborating with the International Agency for Research on Cancer in a study of cancer survival in developing countries. Comparisons with the SEER program results of the National Cancer Institute in the United States, and the EUROCARE study of survival in European countries revealed considerable differences in the survival of patients with certain tumors associated with intensive chemotherapeutic treatment regimes (Hodgkin's disease and testicular tumors), more modest differences in the survival of patients with tumors for which early diagnosis and treatment confer an improved prognosis (carcinomas of the large bowel, breast, and cervix), and only slight differences for tumors associated with poor prognosis (carcinomas of the stomach, pancreas, and lung). With limited resources to meet the challenge of the increasing incidence of cancer expected in the next few decades, health authorities in developing countries should be aware of the importance of investing in a range of cancer control activities, including primary prevention and early detection programs as well as treatment. *Cancer* 1996; 78:2461-4. © 1996 American Cancer Society.

KEYWORDS: survival, cancer, control, developed countries, developing countries.

Most cancers occur in developing countries; 61% of the global incidence in 1985.¹ However, few data are available on cancer incidence, mortality, and, especially, survival in such countries. The reasons for this are clear. Cancer information systems, such as medical records, hospital cancer registries, population-based cancer registries, and mortality registration, are not well established in most in developing countries. Even where some of these are present, difficulties remain in obtaining adequate follow-up information on the vital status of cancer patients. The International Agency for Research on Cancer (IARC) is coordinating a study that aims to provide, for the first time, systematic, population-based information on survival in developing countries, and to compare results obtained in developed countries, especially in Europe and the United States. The IARC study is also providing a context in which to investigate the special problems of cancer registration and patient follow-up, as well as statistical methods for the estimation of survival in developing countries.

See article on page 2581, this issue.

The authors gratefully acknowledge the support of the International Union Against Cancer (UICC), Geneva, Switzerland, the Finnish Cancer Society, Helsinki, Finland, and the Association for International Cancer Research (AICR), St Andrews, Scotland, United Kingdom. Mr. Swaminathan contributed to this work while in receipt of a Special Training Award from the International Agency for Research on Cancer (IARC). They are also thankful to Dr. D. M. Parkin, Chief, Unit of Descriptive Epidemiology, IARC, for his constructive comments on a draft version of this article.

The Study Group on Cancer Survival includes the following members: M. Hamdi-Cherif, Algeria; Jian-Guo Chen, Fan Jin, and Xiang Yong-Bing, China; C. Bratti and A. Rodriguez, Costa Rica; L. Fernandez and M. Graupera, Cuba; R. Sankaranarayanan, R. J. Black, and D. M. Parkin, International Agency for Research on Cancer, France; C. K. Gajalakshmi, K. Jayant, K. Nair, V. Shanta, R. Swaminathan, and B. B. Yeole, India; D. Esteban, the Philippines; N. Martin, S. Sriamporn, and V. Vatanasapt, Thailand.

Address for reprints: R. J. Black, International Agency for Research on Cancer, Unit of Descriptive Epidemiology, 150 cours Albert-Thomas, 69372 Lyon Cedex 08, France.

Received September 26, 1996; accepted October 3, 1996.

TABLE 1
Age-Standardized^a 5-Year Relative Survival of Patients with Selected Cancers (Age, Birth to 74 Yrs)

Site or type of cancer	% of patients surviving				
	U.S. (1967-1973)	U.S. (1974-1986) ^b	U.S. (1986-1991)	Europe (1978-1985)	Developing countries (late 1980s)
Stomach	13	17	20	23	7-17
Large bowel	46	55	60	43	29-37
Pancreas	3 ^c	4	6	6	6-7
Lung	10	15	15	10	3-10
Breast	65	76	82	69	43-63
Cervix	59	68	68	61	27-65
Testis	69	92 ^d	93	85	42-61
Hodgkin's disease	62	77 ^d	79	71	30-55

^a Age standardization involved direct standardization of the site-specific age distributions of the estimated global incidence of major cancers in 1985 (Parkin et al., 1993).¹

^b Includes white patients only.

^c Represents 3-year survival.

^d Represents data for 1981-1986.

RESULTS

The methodology and full results will be published in the IARC Scientific Publications Series. The range of preliminary results obtained from cancer registries in China, Cuba, India, the Philippines, and Thailand are shown in Table 1, alongside previously published data for the U.S.²⁻⁴ and Europe.⁵ Three broad categories of results can be identified. For tumors associated with poor prognosis (stomach, pancreas, and lung), the absolute differences in survival between developed and developing countries were slight. There were greater absolute differences for tumors of the large bowel, breast, and cervix, all of which are associated with a moderate to good prognosis if detected and treated early. For all the tumors in this second group, patient survival was greater in the U.S. than in Europe, at least in the late 1970s to early 1980s, the only period for which comparable data were available. There were also substantial variations among the developing countries themselves, most notably for carcinoma of the breast and cervix. Finally, the greatest differences in survival between the U.S. and developing countries were found for testicular tumors and Hodgkin's disease. Again, there was some evidence of variations among developing countries themselves. Survival of patients with this third group of cancers was only slightly lower in Europe than in the U.S. In general, survival in developing countries was similar to or a little lower than survival in the U.S. in the late 1960s and early 1970s.

INTERPRETATION OF COMPARATIVE SURVIVAL DATA

Population-based survival for patients with a particular cancer represents the "average" outcome achieved in a

population of individuals whose prognosis is influenced by a range of host factors (e.g., age, sex, or risk of death from other diseases), tumor-related factors (e.g., extent of disease), and factors relating to cancer control activities at the population level (e.g., availability and quality of diagnosis and treatment services). Estimates of the survival probability in a group may also be affected by artifacts of case ascertainment by cancer registries, follow-up, and statistical methodology. In comparing results for developed and developing countries, it is important to realize that some or all of these factors may be involved. However, we believe that we have minimized the effects of artifacts in the data by using standardized cancer registration definitions and data collection techniques. Furthermore, some of the host factors could be accounted for by standardizing for age and using relative survival methodology, which adjusts for differences in competing risks of mortality in the populations compared. For tumors associated with poor prognosis (carcinomas of the lung and pancreas), the low rates of survival estimated for developing countries offer some reassurance about the completeness of at least short term follow-up. The remaining differences in survival are likely to be due mainly to the disease being in a later stage at diagnosis and a lack of availability of appropriate treatment for at least some sectors of the populations of developing countries. It should be noted that because the present study is descriptive, the following remarks are somewhat speculative.

The survival of patients with testicular tumors and Hodgkin's disease in developed countries has increased substantially since the introduction of effective chemotherapy and multimodal therapy in the last 25 years. The lack of well-established first-line medical

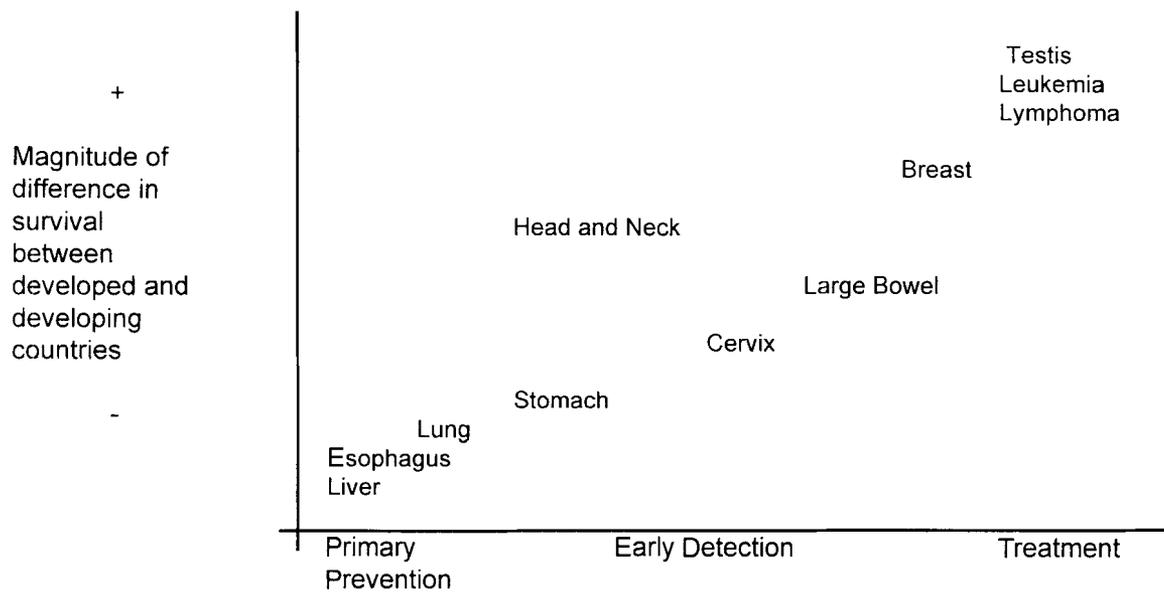


FIGURE 1. Differences in survival of patients with certain cancers between developed and developing countries, in relation to the potential benefits of primary prevention, early detection and treatment, or some combination of these, in reducing cancer mortality in developing countries.

treatment in some developing countries may lead to later diagnosis and referral of patients with these conditions, but it seems likely that the major determinant of survival is the availability of treatment.

For carcinomas of the large bowel, breast, and cervix, the relative importance of the stage of disease at diagnosis and the availability of treatment in developing countries is less evident. The differences in the survival of patients with carcinoma of the large bowel between the U.S. and Europe are probably due to earlier diagnosis in the U.S. through extensive use of fecal occult blood testing and endoscopy, which began in the early 1980s.⁶ With less extensive use of these early detection methods in Europe, at least in the early 1980s, the survival rates achieved were not substantially greater than in some developing countries. This suggests that stage of disease is the most important factor affecting survival. The mainstay of the treatment of colorectal carcinoma is surgical management, and developing countries have at least a basic level of facilities available for surgery.

The 5-year relative survival for women younger than 75 years with breast carcinoma was 43–63% in developing countries, as compared with 65% (in 1967–1973), 76% (in 1974–1986), and 82% (in 1986–1991) in the U.S. There has been considerable progress in the early detection and treatment of breast carcinoma in the last 3 decades. There is a debate about the relative contribution of early detection and treatment to the improvement in survival and to the recently observed reductions

in breast carcinoma mortality in some developed countries. Systematic mammographic screening for breast carcinoma was not practiced in any of the developing countries during the period of data collection. Moreover, neither was opportunistic screening widely practiced in developing countries in contrast to many developed countries in the 1980s. It seems likely that the survival differences are due to both late stage of disease at presentation as well as the availability and quality of adjuvant treatment.

Compared with the survival of patients with some other cancers, the differences in patient survival among the registries of most developing countries, the U.S., and Europe were quite modest for cervical carcinoma. This may be because the key elements of control of this disease, early detection and surgical and radiotherapeutic treatment, are within the technologic scope of the health authorities of most developing countries included in this study. Opportunistic cervical screening was practiced in the registry areas of all developing countries during the period of data collection, giving at least some sectors of those populations the opportunity to be treated when disease was in an early stage. The lack of systematic screening of all women at risk means that the overall distribution of stage of disease for patients is less favorable in developing countries (for example, in India⁷) than in the U.S. and Europe.

CONCLUSIONS

Despite concerns about data quality and comparability, and despite the descriptive nature of this initial

study, we believe that the results are noteworthy and informative. Survival results for developed countries demonstrate what can be achieved. Our study represents a first step toward identifying the elements of cancer control, including primary prevention, early detection initiatives, and treatment, which are most likely to contribute to the reduction of cancer mortality in developing countries.

Figure 1 shows the magnitude of differences in survival for patients with certain common cancers between developing countries and developed countries, plotted against an axis representing the different phases of cancer control (primary prevention; early detection, including screening; and treatment). For tumors associated with poor prognosis, there is a lack of effective screening tests, and treatment has a low success rate even for the small proportion of tumors detected at an early stage. For these, primary prevention is currently the only viable strategy. At the other extreme, the considerable differences in survival observed for patients with carcinoma of the testis and Hodgkin's disease are most likely to be reduced by improved treatment in developing countries. Between the extremes, there are a number of tumors for which there is effective treatment of early disease. For these, some combination of early detection measures allied to treatment services offers the best hope of improving survival in developing countries. However, more information is required to identify the most suitable policies. Two types of studies are required. First, "high-resolution" studies of the diagnosis, treatment, and outcome for samples of patients from population-based cancer registries in developing countries are

needed to identify the subgroups of the populations with poor survival (this is the approach taken in following up the EUROCARE descriptive study of survival in Europe, which showed significant differences among European countries in patient survival of certain major cancers). Second, population-based trials are required to identify technologically and economically viable programs of screening allied to accessible treatment facilities.

REFERENCES

1. Parkin DM, Pisani P, Ferlay J. Estimates of the worldwide incidence of eighteen major cancers in 1985. *Int J Cancer* 1993;54:594-606.
2. Axtell LM, Asire AJ, Myers MH. Cancer patient survival: report No. 5. DHEW Publication No. (NIH) 77-992. Bethesda, MD: U.S. Department of Health, Education and Welfare, Public Health Service, National Institutes of Health, National Cancer Institute, 1976.
3. Ries LAG, Hankey BF, Edwards BK, editors. Cancer statistics review, 1973-87. NIH Publication No. 90-2789. Bethesda, MD: National Cancer Institute, 1990.
4. Kosary CL, Ries LAG, Miller BA, Hankey BF, Harras A, Edwards BK, editors. SEER cancer statistics review, 1973-1992: tables and graphs. NIH Publication No. 96-2789. Bethesda, MD: National Cancer Institute, 1995.
5. Berrino F, Sant M, Verdecchia A, Capocaccia R, Hakulinen T, Esteve J, editors. Cancer survival in Europe. IARC Scientific Publication 132. Lyon: International Agency for Research On Cancer, 1995.
6. Chu KC, Tarone RE, Chow W-H, Hankey BF, Ries LAG. Temporal patterns in colorectal cancer incidence, survival, and mortality from 1950 through 1990. *J Natl Cancer Inst* 1995; 86:997-1006.
7. Nandakumar A, Anantha N, Venugopal TC. Incidence, mortality and survival in cancer of the cervix in Bangalore, India. *Br J Cancer* 1995;71:1348-52.

Chapter 2

Statistical methods for cancer survival analysis

Swaminathan R and Brenner H

Abstract

Adequate and complete follow-up is a prerequisite for the conduct of any survival study. Passive follow-up relies on routine availability of mortality data through unique data linkage possibilities, while active follow-up supplements mortality ascertainment, for which there are a variety of methods. Cox proportional-hazard model was employed to test whether censoring was random in presence of loss to follow-up. Absolute survival probability was estimated by the actuarial method following semi-complete approach for all registries, and the period approach was also used wherever possible. Expected survival probability for registries was estimated from the respective country-, age- and sex-specific abridged life tables. Relative survival, as the ratio of absolute to expected survival, was calculated to exclude the effect arising from different background mortalities. To account for the differences in the age structure of the cancer cases, relative survival was adjusted for age and reported as age-standardized relative survival. Estimated incident cancer cases from less-developed countries together for every classified cancer site served as the standard population. Weights were assigned to individual patients, depending on their age, and standardization was carried out using weighted individual data. Analyses were done using the publicly available macros in SAS software.

Introduction and background

The life table, one of the basic tools in the description of mortality experience of a population, was first developed as early as 1693 by E. Halley in England. It forms the basis for calculation of the life table estimate of the survivor function, which is still widely used today in the analysis of data from epidemiological studies. Information on survival has long been recognized as an important component in monitoring cancer control activities [1]. Like all other health indices, survival statistics are useful primarily as comparative measures. It is these comparisons that help us to suggest possible reasons for the variations and provide targets for improvement and a means of monitoring progress towards them [2]. Survival data obtained from a population-based cancer registry ideally portrays the average outcome of the disease in the pertaining region covered since it is based on an unselected series of incident cancer cases [3].

Follow-up

Adequate and complete follow-up is a prerequisite to conducting a survival study. Lengthy periods of time may be required until the event of interest (any death is the outcome studied in this publication) occurs in all cases studied and maintenance of surveillance on

patients may be extremely difficult. Hence, a closing date for follow-up is typically imposed keeping in mind the adequacy of follow-up information needed to estimate the survival at a specified time. Complete follow-up is deemed to have been achieved when the vital status (alive/dead) at closing date is known for an individual. If not known, then the follow-up is incomplete.

With passive follow-up, information on deaths is routinely received either by-law or via an arrangement with the vital statistics division. Using this procedure, those patients for whom no information of death has been received may be considered to be "alive" until that point of time. The main requirement for this method to work efficiently is that there is a high quality of registration of mortality data and unique data linkage possibilities which ensure the follow-up of cases to be complete with the exception of migration or rare losses. A few of the registries contributing data to this scientific publication have relied almost entirely on this means of obtaining follow-up information.

Active follow-up is necessary in the absence of a reliable health information system, and it may supplement the latter in case of incomplete passive follow-up. Most registries that contributed data to

this scientific publication generally resorted to this method after the routine matching of the incident cancer cases with the available mortality information was completed. The different ways by which this is accomplished are by repeated scrutiny of medical records in hospitals, enquiries with attending physicians, scanning the population registers (city directories), health registers of national health services, health insurance registers, electoral lists, postal/telephone enquiries and visits to the homes of the cases or persons known to them.

Censoring

It is impractical to continue follow-up until all cases under study are dead. With a closing date of follow-up in place, for the subjects who are withdrawn wilfully, drop out or are lost from the study before this date and for those who are still alive at this date, only a lower limit on lifetime is available. This is not to conclude that no information is available on them, but that the information is partial. This unique feature in lifetime data analysis, which occurs when exact lifetimes until death are known for only a portion of the individuals in the study and known to exceed certain values in the remainder, is called "censoring".

When censoring occurs, either due to the termination of study at the closing date which is solely technical or due to loss to follow-up that is 'unrelated' to the outcome studied, e.g. death, it is said to be random or non-informative censoring. When censoring occurs due to loss of follow-up which is 'related' to death, it is known as non-random or informative censoring.

Test for random censoring

Little reliance can be placed on the estimated survival assuming random censoring when the magnitude of loss to follow-up is high. In such instances, it is desirable to investigate deviation from randomness of censoring. In this publication, the Cox proportional-hazard model [4] was used whenever the censoring before closure of study or loss to follow-up exceeded 10% of total cases. For this purpose, the outcome studied is the "loss to follow-up" within a specified time from the index date. Since the survival is estimated at five years for the majority of cancer registries in this publication, the time is fixed as five years. All cases censored before closure of the study and having had a follow-up of less than five years constitute the loss to follow-up group, and the rest of the cases who are either dead or known to be alive on the closing date of follow-up are treated as censored for this analysis to detect the presence of informative censoring. Since the Cox model deals with survival time dynamically, the varying patterns of every loss to follow-up at different intervals on the survival time scale are well accounted for. Based on the general availability, the variables or determinants that are tested for association with loss to follow-up are age at diagnosis, sex and extent of disease. An example of this type of analysis is given in Table 1, where the proportion of patients lost to follow-up ranges between 7–16% among categories of age at diagnosis and 0–27% among categories of extent of disease. A statistically significant differential risk of loss to follow-up is observed. This suggests the presence of non-randomness of loss to follow-up and, therefore,

Table 1. Example of test for randomness of loss to follow-up: Cox proportional-hazards model

Registry	: Mumbai			
Site of cancer	: Female breast			
Period of registration of cases	: 1992–1994			
Period of follow-up	: 1992–1999			
Event studied	: Lost to follow-up before 31 st December 1999 and having a follow-up of <5 years			
% Loss to follow-up	: 10.9%			
Determinants of loss to follow-up	Lost to follow-up		Relative hazard of loss to follow-up [§]	
	Number	%	Hazard ratio	95% CI
Age at diagnosis				
≤ 44 years	53	6.9	1.00	-
45–54	75	10.4	1.63	1.14–2.31*
55–64	89	16.2	2.47	1.76–3.47*
65–74	47	13.7	2.25	1.52–3.33*
75+	9	7.4	1.26	0.62–2.56
Extent of disease				
Localized	129	14.6	1.00	-
Regional	98	8.2	0.54	0.41–0.71*
Distant metastasis	1	0.4	0.05	0.01–0.35*
Unknown	45	26.9	2.40	1.70–3.38*

[§] Each factor is adjusted for the other in the table; CI: Confidence interval; * $p \leq 0.05$.

the survival estimates assuming random censoring should be interpreted with caution.

Actuarial method of estimation of absolute survival probability

It is rare to find a closed group of subjects in a survival study without censoring, except possibly in an artificial situation such as the construction of a life table. The actuarial method of estimating survival probability [5] handles censoring by assuming it to be random. This method involves the construction of a life table that permits the calculation of the cumulative probability of survival at time t_{i+1} from the conditional probabilities of survival during consecutive intervals of follow-up time up to $< t_{i+1}$. This method has been used in this publication to estimate the absolute survival probability. The layout and method of calculation of the elements of a life table are illustrated in Table 2 [6].

For each time period t_i to t_{i+1} , n_i is the number of subjects at risk of outcome at the beginning of the time interval. The number of cases censored during the interval, because they are lost to follow-up or withdrawn alive at the end of the follow-up period, is shown as w_i . The symbol d_i denotes subjects who experienced the outcome during each interval. The effective number of subjects at risk during each interval is calculated as:

$$N_i = n_i - \left(\frac{w_i}{2}\right)$$

In this way, subjects who are alive and at risk of experiencing the outcome during the interval t_i to t_{i+1} , but who are censored at some point of time during the interval, are assumed to have been followed up

for, on average, half of the interval. This actuarial assumption is based on the censorings being independent of the outcome studied (i.e., any death, in this publication). The probability of occurrence of the outcome during the interval is given by

$$q_i = \frac{d_i}{N_i}.$$

The probability of survival during the interval beginning t_i is then calculated as

$$p_i = 1 - q_i$$

from which the cumulative probability of survival up to time t_{i+1} is derived from the product of the p_i 's

$$P_{i+1} = \prod_{j=0}^i p_j$$

This quantity p_{i+1} is often multiplied by 100 to give the "percentage survival" at time t_{i+1} .

Different approaches

There are several approaches to estimating the absolute survival at a given time by varying the registration and follow-up periods of time. These are discussed below and illustrated in Figure 1.

Cohort analysis

The simplest way of computing survival probability is to compute the ratio or percentage of the number of subjects alive at the end of, e.g., 5 years from the index date by the total number of subjects in the study at the beginning of the study, excluding those who did not have a chance to be followed for 5 years

Table 2. Illustration of the layout of the life table and calculation of cumulative survival probability by the actuarial method

Interval	Alive at beginning of interval	Last known alive during interval (censored)	No. of deaths during interval	Effective number at risk	Conditional probability of death	Conditional probability of survival	Cumulative probability of survival (to end of interval)
$t_i - t_{i+1}$	n_i	w_i	d_i	N_i	q_i	p_i	P_{i+1}
0-1	3289	166	365	3206.0	0.114	0.886	0.886
1-2	2758	275	301	2620.5	0.115	0.885	0.784
2-3	2182	37	278	2163.5	0.128	0.872	0.683
3-4	1867	30	191	1852.0	0.103	0.897	0.613
4-5	1646	20	106	1636.0	0.065	0.935	0.573

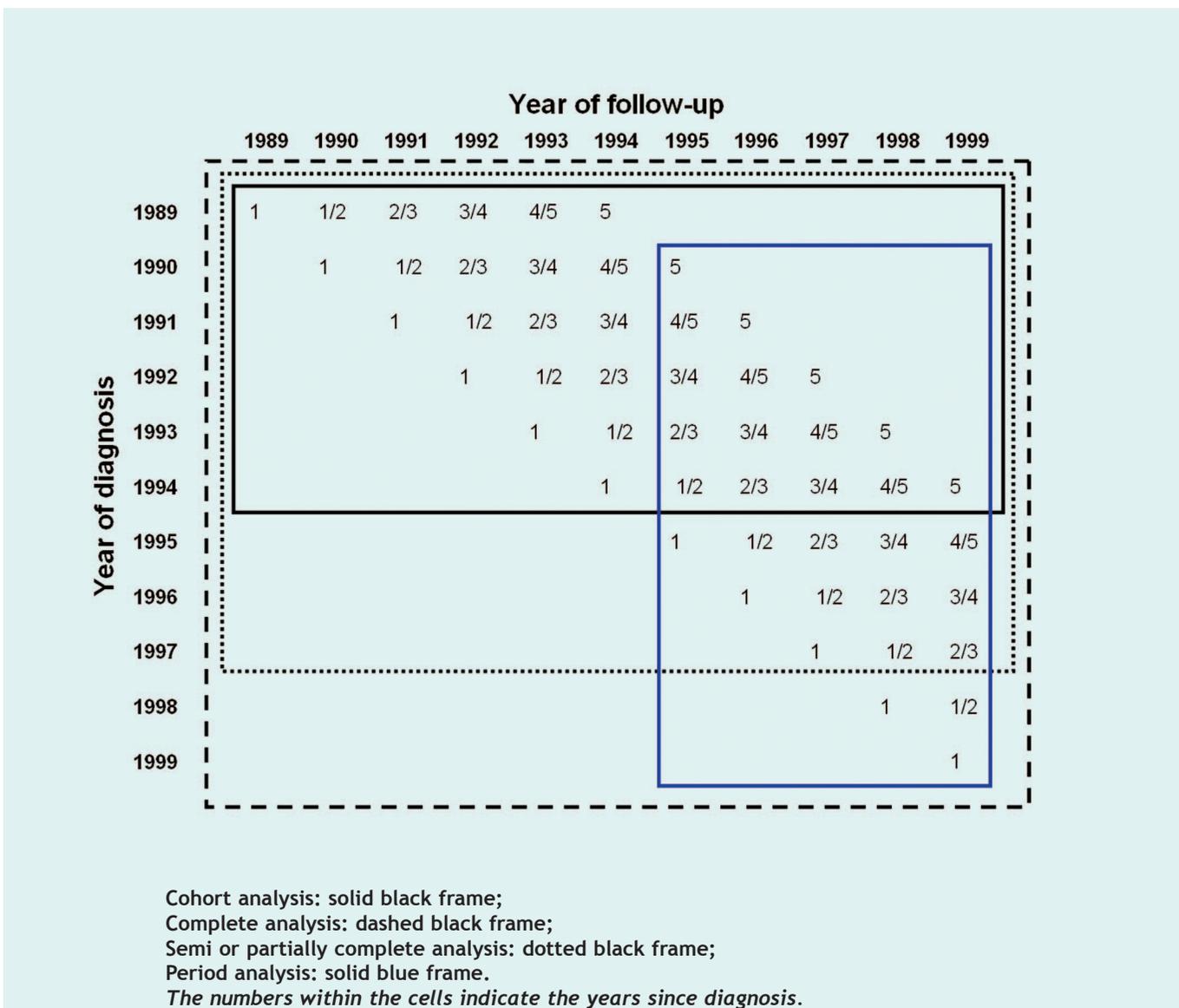
Source: Black and Swaminathan (1998)

after diagnosis. For this purpose, only subjects potentially under observation for at least 5 years and having a potentially complete follow-up of five years are taken into consideration. This approach, which has been called cohort analysis [7] has the disadvantage that even the most recent survival estimates are exclusively based on patients diagnosed many years ago. For example, with a database that includes patients diagnosed between 1989 and 1999 with a closing date of follow-up at the end of 1999, a cohort estimate of 5-year survival could be obtained from patients diagnosed in 1994 at the latest, because patients diagnosed in later years could not possibly have 5-year follow-up by the end of 1999. This approach is illustrated by the solid black frame in Figure 1.

Complete analysis

This is the approach to be used when there is no restriction on the potential follow-up time to equal, e.g., five years from the index date for which the survival is estimated. Rather, all subjects who are diagnosed as incident cancers until the closing date of the follow-up period qualify for inclusion in the analysis. Apart from the subjects with a complete follow-up of five years, those under observation for a variable period of time and having an incomplete follow-up of less than five years are included [7]. In the example given above, all patients diagnosed in 1995–1999 could be included in addition to those diagnosed in earlier years for the derivation of a complete estimate of 5-year survival. This approach is illustrated by the dashed black frame in Figure 1.

Figure 1. Types of analysis to derive up-to-date 5-year survival estimates based on data of patients diagnosed in 1989–1999 and followed until the end of 1999



Semi or partially complete analysis

This approach is widely practised in the estimation of survival by cancer registries. It was adopted in the previous publication on cancer survival [6], and is used for most analyses in this publication as well. Here, not all patients diagnosed until the closing date of follow-up are included. Rather, only patients who have had some minimum potential follow-up time at the closing date of follow-up, such as two or three years, are included. In our example, a partially complete estimate of 5-year survival may be obtained from patients diagnosed in or before 1997 and who have had a minimum of two potential years of follow-up at the end of 1999. This approach, which is in between the pure cohort and pure complete analysis, is illustrated by the dotted black frame in Figure 1.

Period analysis

This is an alternative approach [8] to deriving more up-to-date estimates of cancer patient survival by exclusively utilising the survival information pertaining to the most recent incidence and follow-up periods. The period of interest could be a single calendar year or more. Period analysis exclusively reflects the survival experience of subjects within the most recent calendar period for which the follow-up is available. This is achieved by left truncation of observations at the beginning of this period in addition to censoring at its end [9].

In our example, assume that a period estimate of 5-year survival is to be derived for the 1995–1999 period, the most recent period for which pertinent data are available, then all observations are left truncated at the beginning of 1995 in addition to being censored at the end of 1999. The 5-year period estimate of survival would be obtained from patients diagnosed in 1990–1999 for whom some proportion of 5-year follow-up might have fallen in the 1995–1999 period. With this approach, illustrated by the solid blue frame in Figure 1, different parts of the survival function would be derived from patients diagnosed in various calendar years. Survival during the first year following diagnosis would be estimated for patients diagnosed in 1994–1999, survival during the second year following diagnosis would be estimated for patients diagnosed in 1993–1998, and so on, until survival experience during the fifth year following diagnosis which would be obtained for patients diagnosed in 1990–1995. These conditional survival probabilities are then combined in the usual way to generate 5-year cumulative survival estimates for the 1995–1999 period. It has been shown that period analysis is the approach that clearly provides the most up-to-date estimates of cancer patient survival, and that period estimates of survival for some given

period quite closely predict survival experience of patients diagnosed during that period [10]. In this publication, however, period analysis could not routinely be used because incidence data had not been collected up to the closing date of follow-up by most registries. That said, period analysis was used with data from registries in Qidong and Tianjin, China, and Singapore. A comparison of the survival estimates by cohort and period approaches has been done and the trends over calendar time were depicted.

Relative survival

Berkson [11] in 1942 introduced the concept of relative survival. The relative survival (R_i) for a group of patients at the end of an interval beginning at time t_i is defined as

$$R_i = \frac{S_i}{S_i^*}$$

where S_i is the absolute survival for subjects with a particular cancer and S_i^* is the expected survival of a group of individuals with the same demographic characteristics (age, sex, etc.) who are at risk of death only from causes other than the cancer under study [12]. Berkson and Gage [13] suggested that the observed proportion of survivors of cancer can be compared with an expected proportion of survivors derived from similar people from the general population, most of whom do not have the disease under study. The concept of relative survival methodology has primarily been designed for cancer survival studies to exclude the effect arising from different background mortalities.

Estimation of expected survival probabilities

Expected survival probabilities are usually estimated from age- and sex-specific (sometimes also race-specific) life tables of the general population for the registry area. At least three different methods have been proposed to estimate expected survival, the so-called Ederer I [12], Ederer II [14] and Hakulinen [15] methods. For follow-up times up to 5 years (as reported in this publication) they generally give very similar results. In this study, expected survival probabilities are estimated from country-, age- and sex-specific abridged life tables [16] according to the Ederer II method [14] (for 5-year survival) and the Hakulinen method [15] (for 10- and 15-year survival), the latter of which corrects for potential heterogeneity in patient withdrawal over long potential follow-up times. The estimation of expected survival for earlier calendar periods is done using the country-, age- and sex-specific life tables of the respective calendar periods.

Age-standardization of survival

Most biological phenomena are related to age; there is no reason to expect that survival is not. It is important to note that use of relative rather than absolute survival does not make age-standardization unnecessary. For many types of cancer, the risk of dying as a result of the cancer itself is clearly associated with a subject's age at diagnosis. The ages at diagnosis of cases of any cancer in the developing and developed countries are vastly different [17]. When comparing survival in different groups of patients from different regions, there is a definite need to standardize both absolute and relative survival estimates for age.

For this purpose, direct standardization of survival estimates has been advocated [18]. This is commonly done by using direct standardization of age-specific survival estimates to derive summary statistics called age-standardized absolute survival (ASAS) or age-standardized relative survival (ASRS). For example, ASAS at the end of some follow-up period i is given by

$$ASAS_i = \frac{\sum_x a_{ix} st_x}{\sum_x st_x}$$

where the a_{ix} are age-specific (x :0–4;5–9; etc.) absolute survival estimates at the end of follow-up period t_i and st_x are the age-specific proportions used as "standard or weight" for standardization. The st_x could be arbitrary. Traditionally, the weights have been chosen to reflect the age distribution at diagnosis of some standard cancer population, such as the world standard cancer population [19].

However, for relative survival, the traditional age-standardization, as outlined above, provides results that are conceptually different from crude survival data [20]. Furthermore, traditional age-standardization is often difficult if not impossible to carry out in the presence of sparse and censored data. Hence, in this publication, an alternative approach to age-standardization [21] has been adopted. In this approach, one first assigns the weights to the individual patients depending on their age and then carries out conventional survival analyses using the "weighted individual data". The weights are defined as the ratio of the proportion of patients in the respective age group (x) in the standard population (st) divided by the proportion of patients in the respective age group in the study population. Whereas in the unadjusted (crude analyses), each patient in the study population and her/his contributions to the numbers of persons at risk and deaths are (implicitly) entered with a weight of 1, the proposed form of age-adjustment gives

weights higher (lower) than 1 to patients in age groups which are under-represented (over-represented) in the study population compared to the standard population. The advantages of doing this type of adjustment are: (i) it remains feasible with sparse data, even in situations where survival estimates cannot be derived for certain age groups, and (ii) it provides age-adjusted estimates of relative survival that are conceptually consistent with the crude estimates. In particular, age-adjustment to the study population's own age structure yields a standardized relative survival that is identical to the crude one.

In this study, the weights are defined as the ratio of the proportion of patients in the respective age group in the standard population as summarized in GLOBOCAN 2002 [22], divided by the proportion of patients in the respective age group in the study population registry for every classified cancer site/type.

Software used

While absolute survival can be estimated with any of a large number of commercially available statistical software packages, there are only few specialized programs for relative survival analysis. In this study, analyses are done using the publicly available SAS macros "period" or "periodh" (age-specific and crude analysis, [10]) or "adperiod" or "adperiodh" (age-adjusted analysis, [21]), which can be used to calculate both absolute and relative survival (Ederer and Hakulinen methods) with either the cohort, semi-complete or period approach.

References

1. WHO/IARC. *Cancer Statistics: Report of a WHO/IARC Expert Committee, WHO Technical Report Series No. 632*. World Health Organization, Geneva, 1979.
2. Black RJ, Sankaranarayanan R and Parkin DM. Interpretation of population based cancer survival data. In: *Cancer Survival in Developing Countries*. (eds) R. Sankaranarayanan, RJ Black and DM Parkin. IARC Scientific Publications No. 145. IARC Press, Lyon, 1998, pp 13–17.
3. Sankaranarayanan R, Black RJ and Parkin DM. Eds. *Cancer Survival in Developing Countries*. IARC Scientific Publications No. 145. IARC Press, Lyon, 1998.
4. Cox DR. Regression models and life tables. *J R Statistic Soc (B)* 1972;34: 187–220.

5. Cutler SJ, Ederer F. Maximum utilization of the life table method in analysing survival. *J Chron Dis.* 1958;8: 699–712.
6. Black RJ and Swaminathan R. Statistical methods for the analysis of cancer survival data. In: *Cancer Survival in Developing Countries* (eds) Sankaranarayanan R, Black RJ and Parkin DM. IARC Scientific Publications No. 145. IARCPress, Lyon, 1998; pp 3–7.
7. Brenner H, Gefeller O. Deriving more up-to-date estimates of long term patient survival. *J Clin Epidemiol* 1997; 50: 211–216.
8. Brenner H, Gefeller O. An alternative approach to monitoring cancer patient survival. *Cancer* 1996; 78: 2004–2010.
9. Brenner H, Gefeller O, Hakulinen T. A computer program for period analysis of cancer patient survival. *Eur J Cancer* 2002; 38: 690–695.
10. Brenner H, Hakulinen T. Up-to-date long-term survival curves of patients with cancer by period analysis. *J Clin Oncol* 2002; 20: 826–832.
11. Berkson J. The calculation of survival rates. In: *Carcinoma and other malignant lesions of the stomach.* (eds) W. Wlaters, HK Gray and JT Priestly, 467–484. Philadelphia: Sanders, 1942.
12. Ederer F, Axtell LM and Cutler SJ. The relative survival rate: a statistical methodology. *Monogr Natl Cancer Inst* 1961; 6: 101–121.
13. Berkson J, Gage RP. Calculation of survival rates for cancer. *Proc Mayo Clinic* 1950; 25: 270–286.
14. Ederer F, Heise H. *Instructions to IBM 650 programmers in processing survival computations. Methodological note No. 10.* Bethesda, National Cancer Institute, 1959.
15. Hakulinen T. Cancer survival corrected for heterogeneity in patient withdrawal. *Biometrics* 1982; 38: 933–942.
16. Lopez AD, Ahmad OB, Guillot M, Inoue M, Ferguson BD, Salomon JA. *Life tables for 191 countries for 2000: data, methods, results (GPE discussion paper No. 40). Health Systems Performance Assessments Peer Review, Technical Documentation. IV Outcomes: Population Health. Evidence and information for Policy (EIP).* WHO, October 2001.
17. Parkin DM, Whelan SL, Ferlay J, Teppo L and Thomas DB. Eds. *Cancer Incidence in Five Continents, Vol VIII.* IARC Scientific Publications No. 155. IARCPress, Lyon, 2002.
18. Parkin DM and Hakulinen T. Analysis of survival. In: Jensen OM, Parkin DM, MacLennan R, Muir CS and Skeet RG (eds) *Cancer Registration, Principles and Methods.* IARC Scientific Publications No. 95. IARCPress, Lyon, 1991; pp 159–176.
19. Black RJ and Bashir SA. World standard cancer patient populations: A resource for comparative analysis of survival data. In: *Cancer Survival in Developing Countries.* (eds) Sankaranarayanan R, Black RJ and Parkin DM. IARC Scientific Publications No. 145. IARCPress, Lyon, 1998; pp 9–11.
20. Brenner H, Hakulinen T. On crude and age-adjusted relative survival rates. *J Clin Epidemiol* 2003; 56: 1185–1191.
21. Brenner H, Arndt V, Gefeller O, Hakulinen T. An alternative approach to age-adjustment of cancer survival rates. *Eur J Cancer* 2004; 40: 2317–2322.
22. Ferlay J, Bray F, Pisani P and Parkin DM. *GLOBOCAN 2002: Cancer Incidence, Mortality and Prevalence Worldwide.* IARC CancerBase No. 5, version 2.0. IARCPress, Lyon, 2004.

