Chapter 4

Cancer survival in Africa, Asia, the Caribbean and Central America: Database and attributes

Swaminathan R, Lucas E and Sankaranarayanan R

Abstract

Thirty-one registries in 17 countries submitted data for systematic and centralized scrutiny. Data on 564 606 cases of different cancers ranging 1–56 sites/types from 27 registries in 14 low-/medium-resource countries in Eastern and Western Africa, the Caribbean, Central America and four regions of Asia, registered during 1990–2001 (period varying for individual registries) were reported. The database for this survival study comprised data that were classified as mandatory and optional. Mandatory variables provided by all registries included case-ID, age at diagnosis, sex, incidence date, most valid basis of diagnosis, cancer site/type (ICD-10 codes C00-96), vital status at follow-up and corresponding date. Clinical extent of disease was prominent among the optional variables provided by 17 registries and analysed. The grouping of cancer sites for analysis was based on standard norms, and only categories with at least 25 cases were reported. Cases registered based on a death certificate only, cases lacking any follow-up after initial registration, or cases rejected based on validation checks were excluded from the survival analysis. An easy guide to contents in subsequent chapters, especially tables and graphs describing data quality indices, survival statistics and online dynamic functions, is provided.

Introduction

The database for the survival study was conceived on the basis of data routinely collected in population-based cancer registries in most countries. Accordingly, the variables needed were classified under three major headings: the person, the disease and the follow-up. Under each heading the variables were classified as either mandatory or optional. The variables are summarized in Table 1.

The choice of registries for participation in this study was mainly those whose data on cancer incidence and mortality were published in any volume of *Cancer Incidence in Five Continents* [1]. The response was overwhelming: thirty-one registries in 17 countries submitted data for centralized scrutiny. One or some of the mandatory data variables required were not provided by two of the registries and in four there was a significant incompleteness in follow-up. Hence the data from these four registries were rejected after cursory introspection. Thus data submitted by 27 registries from 14 countries were included for further systematic scrutiny.

Inclusion and exclusion criteria

The broad inclusion and exclusion criteria fixed for this study are given in Table 2. The processing of data for individual registries with a pre-specified set of minimal checks for validity and consistency of data revealed the different procedures followed by some of the registries [2]. The first step undertaken was to standardize the norms to facilitate an unambiguous exclusion of cases from the study. The distinction between a Death Certificate initiated cases (DCN) for further trace-back of information and cases finally registered based on a Death Certificate Only (DCO) without any additional information was established after several exchanges of correspondence with the concerned registries. Such DCO cases with the incidence date the same as the date of death were excluded from survival analysis. In rare instances, the date of follow-up was mistaken to be the date of follow-up attempt rather than the date corresponding to the vital status. Such cases were mostly notified by a code for loss to follow-up or not coded for vital status in the data. These discrepancies were addressed diligently and resolved in consultation with the respective registries to classify them as having incomplete or no follow-up. Cases with lack of any

follow-up were then excluded from the analysis. Cases with multiple primaries were identified both by the registry and by routine checks, and were excluded from survival analysis. Thus, a compact set of validation checks for mandatory and optional data variables, as described in Table 3, was evolved and carried out systematically for all the registries. Cases rejected on the basis of these checks were then excluded from the survival analysis.

Data processing

The data sent by the registries were not in a uniform format. These were all converted as database files (dbf) for uniformity.

Registry code & name

A two-digit code based on the ascending alphabetical order of the participating countries was assigned.

Cancer site or type

The data submitted by the registries did not have a uniform coding format, even for mandatory variables. The most prominent among these was the coding for the primary site and/or histology type of cancer. The calendar period of case registration for this study coincided with the smooth transition in coding practices that most registries were undergoing: from one version of ICD-O to another and thereby to

Table 1. Summary of data variables requested from the registries for the survival study

Person-realated data		Disease-realated data		Follow-up realated data	
Mandatory	**Optional**	**Mandatory**	**Optional**	**Mandatory**	**Optional**
• Case ID • Age at diagnosis • Sex	• Socio-demographic • Socioeconomic	• Incidence date • If ICD-O codes used - Site of primary - Histology type - Behaviour • If ICD-10 codes used - Cancer diagnosis • Basis of diagnosis	• If ICD-10 codes used - Histology type • Tumour grade • Clinical extent of disease • Tumour stage	• Vitas status • Date corresponding to vital status	• None

Note: For analysis of survival trend, mandatory data on person, disease and follow-up related variables for preceding years were requested.

Table 2. Inclusion and exclusion criteria for the survival study

Inclusion criteria	:	A cohort of single primary, incident, invasive cancers diagnosed within a specified consecutive calendar period from 1st January to 31st December with a potential follow-up period of five years or more for a sizeable number of cases
Exclusion criteria	:	Cases registered on the basis of a death certificate notification and continuing to remain as a case of Death Certificate Only (DCO) - Incidence date same as the date of death Cases without any follow-up information after the first registration - Lack of any follow-up Cases rejected on validation checks of submitted variables to the survival database - Details of checks listed separately

Table 3. Details of validation checks carried out and the decision made

Validation checks carried out		Decision made
Age at diagnosis, sex, incidence date and follow-up date unknown	:	Reject case
Multiple cancers signified by duplicate case ID or otherwise	:	Reject case
Duplicate registry ID numbers signifying same case with one cancer	:	Include one case only
Out of range codes		
Age at diagnosis (0–98 years)	:	Reject case
Sex (1–2)	:	Reject case
Incidence and follow-up months (1–12)	:	Reject case
Incidence and follow-up years (as applicable)	:	Reject case
Vital status (1–2)	:	Reject case
ICD-O and ICD-10 codes	:	Reject case
Inconsistent data		
Incidence date > Follow-up date	:	Reject case
Age, site and histology combination (based on IARC CHECK program)	:	Include case if listed as a warning Reject case if listed as invalid
Other logical checks on optional variables submitted for scrutiny	:	As appropriate

ICD-10. Some of them remained with ICD-9 coding even after having changed to higher versions of ICD-O [3–7]. These prompted to have the cancer diagnosis converted into a uniform format and codes of four digits following ICD-10 [8]. The case listings of warnings and invalid conversions were sent to the respective registries, and the queries were resolved by mutual consent. The classification of cancer sites or types was based on the same lines as in *Cancer Incidence in Five Continents*, Volume VIII [9] and is described in Table 4. Only categories with at least 25 cases were considered for analysis and reporting.

Age at diagnosis

This refers to the age in completed years on the incidence date. This was verified with the date of birth when provided. Age unknown cases were excluded, and age above 97 years was coded as 98.

Clinical extent of disease

Though this is an optional variable in this study, it has the greatest significance in correlating local factors with the estimated survival. This data is routinely available or collected by most registries, and in this study seventeen registries submitted this information. 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/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/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.

Index date

The starting date for calculating survival in this study is the incidence date. The definition of incidence date did not reveal any substantial variation between registries. Most of the registries resorted to the first date of unequivocal diagnosis of cancer, by any means, as the incidence date. Other alternatives

encountered were hospital admission date or the date of histological verification. Such a variation might result in minimal differences in short-term survival (say <2 years) and negligent differences for long-term survival [10]. Data on the incidence date was submitted as exact dates or to the level of the month and year of diagnosis. In this study, the index date varied between 1st January 1990 and 31st December 2001, with the period varying for individual registries.

Closing date or date of last follow-up

This date varied between registries and ranged between 31st December 1999 and 31st December 2003. The vital status of each patient was classified as dead, alive or lost to follow-up corresponding to this date. This information was submitted as exact dates or to the level of month and year of follow-up by the registries. To measure extent of incompleteness in follow-up, especially for registries that employed active methods of follow-up, a variable called 'Follow-up' was created, and the extent of loss to follow-up in years from index date was classified as <1 year, 1–3 years, 3–5 years and >5 years on survival time.

Survival time

This was calculated as the time (in months) between the index date and the date of death from any cause, date of loss to follow-up or the closing date, whichever was earliest.

Most valid basis of diagnosis

The codes for the most valid basis of diagnosis were also different among registries. Based on the key to these codes, a new variable, 'Histological verification', was created for unambiguity and uniformity.

Inclusion status

Systematic validation of the data was undertaken for the registries by performing the checks listed in Table 3. More customized checks were performed depending on the data type on optional variables provided by each registry. A list of potential errors was returned to the registries for clarification. Registries undertaking follow-up predominantly by passive methods were urged to improve follow-up by resorting to feasible active methods like repeated scrutiny of medical records and linkage of data at sources of registration of cases. After the rectification of errors, if any, the checks were repeated on the revised data. The inclusion status was then classified as follows: Included (I) or excluded for reasons of being a DCO case (D), with lack of any follow-up (F) or due to any other reasons (O) on validation checks.

Data quality indicators

The indices that would determine the data quality can be summarized as follows:

- The frequency of cases, expressed as number and percentage, that were registered as a DCO;
- The frequency of cases, expressed as percentage, that had a histologically confirmed cancer diagnosis;
- The frequency of cases, expressed as number and percentage, that were excluded from survival analysis including those with lack of any follow-up or other errors;
- The frequency of cases, expressed as number and percentage, with incomplete follow-up.

All of the above, by classified cancer site or type, are included as standard tables in the chapters dealing with individual registry data.

Study database

Two databases were created for analysis and reporting of results:

- The file SURVDB.DBF deals with all cases submitted for scrutiny and includes 16 variables. This file is essentially for eliciting the data quality.
- The file SURVDB2.DBF deals with cases included for analysis and comprises 10 variables. This file is essentially for eliciting data on survival.

The description of the variables is given in Table 5.

A lead to the chapters on individual registries

An overview of the issues in the background of the survival data and the analysis carried out for each registry is given in Table 6, as a lead to the forthcoming chapters on individual registries. The cancer registration and follow-up are seen to be completely carried out by active methods in eleven registries (all five from India, two from Thailand, one

Table 4. Details of validation checks carried out and the decision made

ICD-10 code	ICD-10 title	ICD-10 description
C00	Lip	Lip
C01–02	Tongue	Base, other and unspecified tongue
C03–06	Oral cavity	Gum, floor of mouth, palate, buccal mucosa and other mouth
C07–08	Salivary gland	Major salivary glands
C09	Tonsil	Tonsil
C10	Other oropharynx	Vallecula, anterior surface of epiglottis and other oropharynx
C11	Nasopharynx	Nasopharynx
C12–13	Hypopharynx	Pyriform fossa, post cricoid and other hypopharynx
C15	Oesophagus	Oesophagus
C16	Stomach	Stomach
C17	Small intestine	Small intestine
C18	Colon	Colon
C19–20	Rectum	Rectosigmoid junction and rectum
C21	Anus	Anal canal and anus
C22	Liver	Liver
C23–24	Gallbladder	Gallbladder and unspecified biliary tract
C25	Pancreas	Pancreas
C26	Gastrointestinal tract unspecified	Other unspecified gastrointestinal tract
C30–31	Nose/Sinuses	Nasal cavity, middle ear and other accessory sinuses
C32	Larynx	Larynx
C33–34	Lung	Trachea, bronchus and lung
C37–38	Other thoracic organs	Thymus, mediastinum, pleura and heart
C40–41	Bone	Bone, joints and articular cartilage
C43	Melanoma skin	Melanoma of skin
C44	Other skin	Non-melanoma of skin
C45	Mesothelioma	Mesothelioma
C46	Kaposi sarcoma	Kaposi sarcoma
C47;C49	Connective tissue	Peripheral, connective and other soft tissues
C48	Peritoneum	Retroperitoneum and peritoneum
C50	Breast	Male and female breast
C51	Vulva	Vulva
C52	Vagina	Vagina
C53	Cervix	Cervix uteri
C54	Corpus uteri	Endometrium and corpus uteri
C55	Uterus unspecified	Unspecified uterus
C56	Ovary	Ovary
C57	Other female genital	Other and unspecified female genital organs
C58	Placenta	Placenta
C60	Penis	Penis
C61	Prostate	Prostate
C62	Testis	Testis
C63	Other male genital organs	Other and unspecified male genital organs
C64	Kidney	Kidney
C65	Renal pelvis	Renal pelvis
C66	Ureter	Ureter
C67	Urinary bladder	Urinary bladder
C68	Other urinary organs	Other urinary organs
C69	Eye	Eye and adnexa
C70–72	Brain & nervous system	Meninges, brain and other parts of central nervous system
C73	Thyroid	Thyroid gland
C74	Adrenal gland	Adrenal gland
C75	Other endocrine	Other endocrine glands and related structures
C81	Hodgkin lymphoma	Hodgkin lymphoma
C82–85;C96	Non-Hodgkin lymphoma	Non-Hodgkin lymphoma
C90	Multiple myeloma	Multiple myeloma
C91	Lymphoid leukaemia	Lymphoid leukaemia
C92–94	Myeloid leukaemia	Myeloid, monocytic and myeloblastic leukaemia
C95	Leukaemia unspecified	Unspecifed leukaemia

Table 5. Cancer Survival in Africa, Asia, the Caribbean and Central America, database

Database format for all cases submitted (SURVDB.DBF)			Database format for all cases included for survival analysis (SURVDB2.DBF)		
Variable name	**Codes/Description**	**Field length**	**Variable name**	**Codes/ Description**	**Field length**
Registry code	Codes to identify the registry	2	Registry code	Codes to identify the registry	2
Registry name	Name of the registry & country	25	Cancer diagnosis	2 digit ICD-10 codes (numeric only)	2
Cancer diagnosis	4 digit ICD-10 codes	4	Age at diagnosis	00–98 in completed years 98 for 98+ years of age	2
Age at diagnosis	00–98 in completed years 98 for 98+ years of age	2	Sex	1: Male; 2: Female	1
Sex	1: Male; 2: Female	1	Clinical extent of disease	1: Localized; 2: Regional; 3: Distant metastasis; 4: Unknown	1
Clinical extent of disease	1: Localized; 2: Regional; 3: Distant metastasis; 4: Unknown	1	Year of diagnosis	As applicable	4
Histology	ICD-O codes including type, behaviour and grade of tumour	6	Month of diagnosis	01–12	2
			Year of follow-up	As applicable	4
Year of diagnosis	As applicable	4	Month of follow-up	01–12	2
Month of diagnosis	01–12	2	Vital status	1: Alive; 2: Dead	1
Year of follow-up	As applicable	4			
Month of follow-up	01–12	2			
Vital status	1: Alive; 2: Dead	1			
Inclusion status	I: Included; D: DCO; F: Lack of follow-up; O: Excluded for other invalid reasons	1			
Histological verification	0: No; 1: yes	1			
Type of follow-up	C: Complete follow-up at 5-years; 1: Lost to follow-up (LFU) within 1-year; 3: LFU between 1-3 years; 5: LFU between 3-5 years; 6: LFU after 5 years	1			
Survival time	In months	3.1			

each from Pakistan, Turkey, Uganda and Zimbabwe). Among the remaining four registries wherein cancer registration is carried out entirely by active methods, the follow-up methods are done in a mixed manner: predominantly by active methods with a minimal passive component in Manila and Rizal, Philippines, and predominantly by passive methods in Tianjin, China and Incheon, South Korea. Cancer registration and follow-up are entirely done by passive methods in only Singapore. In Hong Kong, the cancer registration

Table 6. An overview of basic characteristics of survival data and analysis done by registry

Cancer Survival in Africa, Asia, the Caribbean and Central America

Country/registry	Method adopted for		Estimation of survival probability by			Analysis of survival	
	Cancer registration	Follow-up	Cohort analysis	Semi complete analysis	Period analysis	Trend	Extent disease
CHINA							
Hong Kong SAR	P + A	P		Y			Y[$]
Qidong	P + A	A + P	Y	Y*	Y	Y	
Shanghai	P	P + A		Y		Y	
Tianjin	A	P + A	Y	Y	Y	Y	
COSTA RICA	P	P + A		Y			Y
CUBA	P	P + A		Y		Y	Y
The GAMBIA	P + A	A		Y			
INDIA							
Barshi	A	A		Y		Y	
Bhopal	A	A		Y			Y
Chennai	A	A		Y		Y	Y
Karunagappally	A	A		Y			Y
Mumbai	A	A		Y		Y	Y
PAKISTAN							
South Karachi	A	A		Y			Y
PHILIPPINES							
Manila*	A	A+ P		Y			Y
Rizal	A	A + P		Y		Y	Y
REPUBLIC OF KOREA							
Busan	P + A	P + A		Y			
Incheon	A	P + A		Y			
Seoul	P + A	P + A		Y			
SAUDI ARABIA							
Riyadh	P + A	P + A		Y			Y
SINGAPORE	P	P	Y	Y	Y	Y	Y
THAILAND							
Chiang Mai	A	A		Y		Y	Y
Khon Kaen	A + P	A + P		Y		Y	Y
Lampang	P	P + A		Y			Y
Songkhla	A	A		Y			Y
TURKEY							
Izmir	A	A		Y			Y
UGANDA							
Kampala	A	A		Y			
ZIMBABWE							
Harare	A	A		Y			

A: Active; P: Passive; A + P: Predominantly active; P + A; Predominantly passive; Y: Yes
** Complete analysis; [$] for breast cancer only*

is mixed while the follow-up is entirely by passive methods. The cancer registration and follow-up are both carried out by a mixture of active and passive methods in all the other registries.

The various approaches to estimating survival probability are illustrated in Chapter 2. The analysis by the semi-complete approach has been done for all the registries excepting Qidong, for which the analysis was done by the complete approach. The analysis of survival trend by the semi-complete approach involving two calendar periods was done for eleven registries, while a comparative analysis of survival trend by cohort and period approaches for several calendar periods was possible in Qidong and Tianjin, China and Singapore. Analysis of survival by clinical extent of disease for selected cancer sites was possible in 17 registries.

A guide to the tables and graphs in the individual registry chapters

A chapter is dedicated to each participating registry. It comprises a concise summary describing the background and salient features of the results combined with standard/optional tables and figures.

Table 1 deals with the main data quality indices prior to and after the commencement of follow-up. It gives the total number of cases registered, proportion (%) of histologically verified diagnosis, frequency of type of exclusions from study like DCOs, lack of follow-up and others, and the total number of excluded and included cases for the study for each classified cancer site/type category.

Table 2 refers to the data quality index on completeness of follow-up. For registries that resorted to passive means of follow-up entirely, this table gives the distribution of vital status (alive/dead) by classified cancer site/type. For others, this table gives the frequency of cases with complete follow-up at the closing date, as well as at 5 years from the index date, and the extent of incompleteness in follow-up by duration (classified years from diagnosis) of loss to follow-up for every classified cancer site/type. The non-randomness of loss to follow-up or informative censoring is indicated wherever encountered. The median follow-up (in months) is given for every registry and classified cancer site/type.

Table 3 gives the crude and age-adjusted survival statistics. Absolute and relative survival (%) at one, three and five years from index date and 5-year age-standardized relative survival for all ages together and for the age interval 0-74 years are given for every classified cancer site/type.

Table 4 deals with survival statistics by sex and classified age groups. The frequency of cases and five-year absolute and relative survival (%) by sex and the frequency of cases and five-year relative survival (%) by the age groups 0–44, 45–54, 55–64, 65–74 and 75+ years are given for every classified cancer site/type.

Figure 1a portrays the top five or ten cancers ranked by 5-year relative survival for those registries that contributed data on sufficient number of cancer sites/types.

Figure 1b displays the top five cancers ranked by 5-year relative survival among males.

Figure 1c represents the top five cancers ranked by 5-year relative survival among females.

Analysis of survival by clinical extent of disease (Table 5)

This is carried out for selected cancer sites only: cancers of the head and neck, female breast, cervix and ovary. The frequency (%) of cases by classified clinical extent of disease categories and the corresponding 5-year absolute survival are presented as a table.

Figure 2 either depicts the absolute survival by clinical extent of disease for available cancer sites and registries or trend of survival by cohort and period approaches as appropriate.

Analysis of trend of survival (Table 6)

This is done in two ways:

For registries that provided data for any one preceding calendar period of time, the 5-year absolute and relative survival were estimated by semi-complete approach for the two periods for the available cancer sites/types and presented as a table.

For registries that provided data for more than two 5-year calendar periods preceding the latest one, the 5-year absolute and relative survival were estimated for the latest two calendar periods for the available cancer sites/types and presented as a table. Additionally, the frequency of cases by 5-year calendar periods by cancer site/type is given in a table. Correspondingly, the five-, ten- and fifteen-year relative survival were estimated by cohort and period approaches for available cancer site/type depending on the availability of data and presented as one or two tables, as necessary. These are depicted as figures also.

Online features of the publication

A dedicated website has been designed to host the new version of Cancer Survival in Africa, Asia, the Caribbean and Central America (SurvCan), available at http://survcan.iarc.fr. Users will be able to access all the chapters of the publication, including abstracts, tables and figures, and will be able to export each chapter in full or part in PDF format. Users will also be able to access the previous volume of the publication (1998) if available. References cited in the chapters are directly linked to PubMed or to the specific website of the publication as appropriate. Online dynamic functions are also supplied to let users generate comparative statistics: users will be able to list all the available tables/figures for a registry, compare values between registries for a specific cancer site, based on ICD-10 codes, and generate specific dynamic figures on survival statistics (5-year absolute survival or 5-year relative survival, etc.) according to the sex, age group and extent of disease. An online help tool is available to facilitate the use of the online statistical functions.

References

1. Parkin DM, Whelan SL, Ferlay J and Storm H. *Cancer Incidence in Five Continents, vol I to VIII. IARC Cancerbase No 7*. IARCPress, Lyon, 2005.

2. Swaminathan R, Black RJ and Sankaranarayanan R. Database on Cancer Survival from Developing Countries. In: *Cancer Survival in Developing Countries* (eds) R.Sankaranarayanan, RJ Black and DM Parkin. IARC Scientific Publications No. 145. IARCPress, Lyon, 1998.

3. *WHO. International Classification of Diseases for Oncology (ICD-O), First edition*. World Health Organization, Geneva, 1976.

4. *WHO. International Classification of Diseases for Oncology (ICD-O), Second edition*. World Health Organization, Geneva, 1990.

5. *WHO. International Classification of Diseases for Oncology (ICD-O), Third edition*. World Health Organization, Geneva, 2000.

6. *WHO. International Classification of Diseases, Ninth Revision* (ICD-9). World Health Organization, Geneva, 1976.

7. *WHO. International Statistical Classification of Diseases and Related Health Problems, Tenth Revision* (ICD-10), Volume 1. World Health Organization, Geneva, 1992.

8. Ferlay J. *IARCcrgTools, Version 1.01*. IARCPress, Lyon, 2003.

9. Parkin DM, Whelan SL, Ferlay J, Teppo L and Thomas DB. *Cancer Incidence in Five Continents, vol VIII*: IARC Scientific publications No 155. IARCPress, Lyon, 2002.

10. Berrino F, Sant M, Verdecchia A, Capocaccia R, Hakulinen T and Esteve J. (eds) *Survival of Cancer Patients in Europe: the EUROCARE Study. IARC Scientific Publications No. 132*. IARCPress, Lyon, 1995.