



LING YANG

Estimating Cancer Burden in China



ACADEMIC DISSERTATION

To be presented, with the permission of
the Faculty of Medicine of the University of Tampere,
for public discussion in the auditorium of Tampere School of
Public Health, Medisiinarinkatu 3, Tampere,
on December 16th, 2005, at 12 o'clock.

Acta Universitatis Tampereensis 1122

ACADEMIC DISSERTATION

University of Tampere, Tampere School of Public Health
Finland

Supervised by
Professor Anssi Auvinen
University of Tampere
Dr. Donald Maxwell Parkin
University of Oxford, UK

Reviewed by
Professor Timo Hakulinen
Helsinki, Finland
Professor Alan Lopez
University of Queensland, Australia

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

Tel. +358 3 3551 6055
Fax +358 3 3551 7685
taju@uta.fi
www.uta.fi/taju
<http://granum.uta.fi>

Cover design by
Juha Siro

Layout
Marita Hallila

Printed dissertation
Acta Universitatis Tamperensis 1122
ISBN 951-44-6474-5
ISSN 1455-1616

Electronic dissertation
Acta Electronica Universitatis Tamperensis 489
ISBN 951-44-6475-3
ISSN 1456-954X
<http://acta.uta.fi>

Tampereen Yliopistopaino Oy – Juvenes Print
Tampere 2005

Contents

1. Summary	5
2. List of Original Communications	7
3. List of abbreviations	8
4. Introduction	9
5. Review of the Literature.....	11
5.1 Resources for cancer information	11
5.1.1 Health information system.....	11
5.1.2 Health Surveillance System	12
5.1.3 Cancer Surveillance System	13
5.1.4 Major components of a cancer surveillance system.....	14
5.2 Sources of information about cancer statistics in the world.....	17
5.2.1 The CancerMondial Project.....	17
5.2.2 Other cancer statistics in regions, areas or countries.....	19
5.3 Sources of information on cancer statistics in China	25
5.3.1 Mortality data from two national retrospective death surveys.....	26
5.3.2 Mortality data from routine reporting systems or special projects	31
5.3.3 Incidence and mortality data from cancer registries	33
5.4 Statistical indices and methods used in cancer statistics.....	35
5.4.1 Measurements of cancer burden.....	35
5.4.2 Comparing cancer burden between different populations	41
5.4.3 Estimation of cancer statistics based on limited data	42
5.4.3.1 Methodological approaches for estimating cancer statistics.....	42
5.4.3.2 Cancer estimation in practice	45
5.4.4 Trends in incidence and mortality	46
5.4.5 Projection of cancer statistics.....	51
6. Aims of the Study	55
7. Materials and Methods.....	56
7.1 Data sources.....	56
7.2 Methods used in the study	56
8. Results.....	61
8.1 Estimated population by area for year 2005 in China.....	61
8.2 Cancer registration practice in China	62
8.3 Validity of cancer mortality data from available data sources in China	64
8.4 Cancer mortality and trends during 1987–1999 in China	66
8.5 Cancer mortality and incidence estimates and projection for 2000 and 2005.....	70
8.5.1 Sensitivity analysis.....	70
8.5.2 Estimation and projection of cancer patterns in 2000 and 2005	71

9. Discussion	77
9.1 Considerations of information sources.....	77
9.2 Methodological considerations.....	80
9.3 Determinants of changing cancer burden in China.....	82
9.4 Challenges for cancer registration in China	87
9.5 Prospects for cancer control in China	88
10. Conclusions	92
11. Acknowledgements.....	94
12. References	96
13. Original Publications.....	108

1. Summary

The aim of the present study was to estimate and project the cancer burden at the national level in China in recent years, from both incidence and mortality profile, in order to provide background information relevant to planning for cancer prevention and for a control programme.

A first national survey on cancer registration practices in China was conducted to identify all, or as many as possible, of the existing cancer registries in the country, to describe the methods being used, and to identify future needs with respect to technical support. A wide variation in practices between the 48 cancer registries in China was found in the survey. Lack of qualified personnel, insufficient funding support and lack of stability of the population were identified as the major problems in registration practice in China. Several suggestions are made about ways to improve availability and quality of incidence and survival data from cancer registries.

After reviewing all currently available data sources used in evaluating the cancer burden in China, a comparison study for the validity and representativity of mortality data from available sources in China was carried out. None of the most available sources cover a random sample of the Chinese population, with respect to age group, sex, and urban-rural residence. However, the sex and region (urban/rural) specific age-standardised mortality rates from the dataset from the Center of Health Information and Statistics (CHIS) of the Ministry of Health were closer to the results of the national mortality survey in 1990–1992, than those calculated from the “Disease Surveillance Points” (DSP) survey data.

Based on the CHIS mortality data, using the Joinpoint regression model, time trends in the period 1987–1999 for nine major cancers were analysed. During the study period, age-standardized mortality rates for cancers of the oesophagus, stomach, cervix uteri, leukaemia (except for urban males after 1996), and nasopharynx declined, while lung cancer and female breast cancer showed significant increasing trends in both urban and rural areas and for both sexes. Cancers of the colon-rectum and liver had different trends in mortality in urban and rural populations. Examining age-specific trends in mortality rates reveals trends in younger generations that differ from the overall (all age) trends which may presage future overall trends.

Starting from the national mortality data of 1991, following the annual rates of change for major cancers, by age, sex and areas from the CHIS data during the period 1991–1999, based on a log-linear regression model with Poisson errors, the national

cancer mortality in the years 2000 and 2005 were estimated and projected. Compared with the second national mortality survey done in 1990–1992, age standardized mortality rates for all cancers combined in 2005 are lightly changed (–0.8% in men and +2.5% in women), but population growth and ageing will result in an increasing number of deaths, from 1.2 million to 1.8 million a year. The largest predicted increases are for the number of female breast cancers (+155.4%) and lung cancers (+112.1% in men, +153.5% in women).

Using ratios of cancer incidence to mortality from those cancer registries matching the quality criteria of Cancer Incidence in Five Continents, 8th version, together with the above national mortality data for year 2000 and 2005, the cancer incidence at the national level was evaluated. A total of 2.1 million cancer cases are estimated for the year 2000 (1.3 million in men, 0.8 million in women), with the most common sites being lung, liver and stomach in men, and breast, lung and stomach in women. The rising rates of lung cancer incidence (in both sexes) and breast cancer mean that there will be much greater increases in the numbers of cases at these two sites (27% for lung cancer in men, 38% for lung and breast cancer in women). These two cancers are now the priorities for cancer prevention, early detection and therapy in China.

Different statistical methods or models (Joinpoint, Poisson regression, log-linear regression) were used and compared with other models shown in sensitivity analyses. The interpretation of the changes of the mortality and incidence profile in China falls into three aspects: demographic changes (population growth and ageing), changes in cancer risks due to such factors as lifestyle, dietary, socioeconomic and environmental change, and the effects of early detection and treatment mediated via governmental health care programmes, policies and resource management. Along with the big improvements in social and economic conditions, the environment, living conditions and lifestyle characteristics are changing rapidly in China, as well as the impressive improvements in the organisation and provision of health care, facilities and human resources. The improvements in socioeconomic status, diet and nutrition may be responsible for the declining risk of some cancers (oesophagus, stomach, nasopharynx), while increasing the risk for others (breast, colon-rectum). Screening programmes (especially for cervical cancer), and more widely available and better facilities for cancer therapy may have helped to reduce mortality and incidence for several cancers. However, tobacco smoking remains a major problem and influences many cancer trends. The large increases in absolute number of new cancer cases and deaths resulting from population increase and aging are much more important in determining the future cancer burden than any other changes due to risk factors, emphasizing the increasing importance of cancer as a health problem in 21st century in China.

2. List of Original Communications

This thesis is based on the following original articles. Some previously unpublished data and results not reported in the original publications are also presented.

- I. Yang L, Parkin DM, Li LD and Chen YD (2004a) Time trends in cancer mortality in China: 1987–1999. *Int J Cancer* 106: 771–783
- II. Yang L, Parkin DM, Li LD and Chen YD (2004b). A comparison of the sources of cancer mortality in China. *Cancer Cause Control* 15: 681–687
- III. Yang L, Parkin DM, Li LD, Chen YD and Bray FI (2004c) Estimation and projection of the national profile of cancer mortality in China: 1991–2005. *Brit J Cancer* 90: 2157–2166.
- IV. Yang L, Parkin DM, Ferlay J, Li LD and Chen YD (2005a) Estimates of cancer incidence in China for 2000 and projection for 2005. *Cancer Epidemiol Biomarkers Prev* 14: 243–250
- V. Yang L, Parkin DM, Whelan S, Zhang SW, Chen YH, Lu FZ and Li LD (2005b) Statistics on cancer in China – cancer registration in 2002. *Eur J Cancer Prev* 14: 329–335

3. List of abbreviations

ANCR: Association of the Nordic Cancer Registries
ACS: American Cancer Society
CAMS: Chinese Academy of Medical Sciences
CAPM: Chinese Academy of Preventive Medicine
CDC: Centers for Disease Control and Prevention
CHIS: Center of Health Information and Statistics
CI5: Cancer Incidence in Five Continents
CINA: Cancer in North America
CSR: Cancer Statistics Review
CTSU: Clinical Trial Service Unit
DALYs: Disability-Adjusted Life Years
DCO: Death Certificate Only
DSP: Disease Surveillance Points
EAPC: Estimated Annual Percent Change
ENCR: European Network of Cancer Registries
EU: European Union
EUROCIM: European Cancer Incidence and Mortality Database
GBD: Global Burden of Disease
HIS: Health Information System
HSS: Health Surveillance System
IACR: International Association of Cancer Registries
IARC: International Agency for Research on Cancer
IBM: Incidence-Based Mortality
ICD: Classification of Diseases
ICIDH: International Classification of Impairments, Disabilities and Handicaps
MOH: Ministry of Health in China
NAACCR: North American Association of Central Cancer Registries
NCCP: National Cancer Control Programmes
NCCR: National Center of Cancer Registries
NCD: Noncommunicable diseases
NCHIS: National Center for Health Statistics Information
NCHS: National Center for Health Statistics
NCI: National Cancer Institute
NCSCH: National Cancer Statistics Clearing House
NOCP: National Office for Cancer Prevention and Control
NVSS: National Vital Statistics System
PBCR: Population-Based Cancer Registry
PYLLs: Person-Years of Life Lost
QALYs: Quality-Adjusted Life Years
SEER: Surveillance, Epidemiology, and End Results
SRR: Standardized Rate Ratio
STEPS: Stepwise Approach to Surveillance
SVR: Sample Vital Registration
VA: Verbal Autopsy
WHO: World Health Organization
WHOSIS: WHO Statistical Information System

4. Introduction

Noncommunicable diseases (NCD), especially cancer, cardiovascular diseases and diabetes are major public health issues in almost all countries in the Asian Pacific Region (WHO 2005a). Driven principally by world population growth and aging, cancer is becoming a major health problem for most countries. According to the estimation from the International Agency for Research on Cancer, in 1990, the estimated global numbers were 8.1 million new cancer cases (Parkin et al. 1999) and 5.2 million deaths (Pisani et al. 1999), while in 2000, 6.2 million died from malignant tumours, responsible for 12% of the nearly 56 million deaths worldwide from all causes. The estimated number of new cases each year is expected rise from 10 million in 2000 (5.3 million men and 4.7 million women) to 15 million by 2020. Some 60% of all these new cases will occur in the less developed parts of the world, and cancer is emerging as a major public health problem in developing countries, matching its effect in industrialized nations. (Parkin 2001, Stewart and Kleihues 2003)

With this increasing burden, understanding, preventing and controlling malignant neoplasm is an urgent priority worldwide. In 2002, the World Health Organization (WHO) published 'National Cancer Control Programmes (NCCP), policy and managerial guidelines' which offers the most rational means of achieving a substantial degree of cancer control, even where resources are severely limited (WHO 2002a). NCCP is a public health programme aiming to reduce cancer incidence and mortality and improve quality of life of cancer patients through the systematic and equitable implementation of evidence-based strategies for prevention, early detection, diagnosis, treatment, and palliation while making the best use of available resources. It is based on current evidence suggesting that at least one-third of the new cases of cancer each year throughout the world are preventable by modifying risk factors (such as controlling tobacco and alcohol use, moderating diet, and immunizing against viral hepatitis B); early detection and effective treatment would permit a further 1/3 of deaths to be avoided where resources are available; while effective techniques permitting comprehensive pain relief and palliative care for improving the quality of life of the one third more advanced cases (and their families) (WHO 2002a). Establishing a comprehensive NCCP requires competent management and the best use of available resources for planning, implementing and evaluating disease control strategies, tailored to the local socioeconomic and culture context, as well as scientific knowledge and

experience ranging from the complexities of intracellular molecular regulation to individual lifestyle choices.

First of all, when setting up an NCCP, a basic information platform is needed for monitoring the programme processes and indicating specific changes related to cancer occurrence and outcome. This information system can evaluate the cancer burden in the country and identifying the major problems which, when combined with the social-economic and health resources would define a NCCP's development tailored to the local context. The resources available are mainly from the health surveillance system based on cancer registries or other statistical sources, and ad hoc surveys. The estimation of national cancer burden and trends depends upon the availability of the accurate cancer statistics, and this is an essential component of the planning and monitoring the cancer control activities. Furthermore, at national level, accurate data are important for planning cancer control activities (such as cancer prevention, screening and treatment) and for monitoring the effects of such interventions.

In China, with one fifth of the world's population, and one manifesting continuing growth and aging, cancer is becoming a major health problem. Although China has an exceptionally long history of civil registration, no national level mortality data are available regarding the national cancer burden. According to the estimate by IARC, which is based on information from vital statistics and cancer registries in the country, a large fraction of the global cancer burden occurs in China (Parkin 2001). Therefore, inferences about global cancer burden, trends and the overall effectiveness of cancer control worldwide are heavily influenced by China, and accurate estimates of incidence and mortality are therefore very important both in a global context and for national cancer control programme setting, as well as providing epidemiological and etiological clues for cancer research in a population with a huge variation in social, economic and health status. It is also recommended that health statistics should be periodically assessed to ensure that they meet the needs of the local health system.

5. Review of the Literature

5.1 Resources for cancer information

5.1.1 *Health information system*

One of the objectives of WHO is to support countries in their efforts to develop health systems that ensure the delivery of effective health services and care for the entire population, based on the primary health care approach. The production and dissemination of health statistics for health action at country, regional and global levels is a core WHO activity mandated to WHO by its member states in its constitution. The multiple roles emphasized by the Organization, including advocacy for health issues, monitoring and evaluation of health programmes, provision of technical assistance to countries require the best possible health statistics from the countries covered. Therefore, improving WHO's work in health information becomes an emphasized integrated project. (WHO 2000)

Five steps need to be taken as a problem-solving chain in the formation of health strategies in populations: definition, identification and description of disease or other health-related phenomenon, definition of target and target groups, setting of priorities, decision and action, evaluation (Eylenbosch and Noah 1988). Information concerning the health of the population is needed at all stages of the problem-solving process to describe the burden, causes and consequences of the disease, and to establish the necessary platform for priority setting concerning objective and for the choice of target groups. For the decision-makers and planners, the surveillance and monitoring system will be tools for controlling what happens to the population's health.

A Health Information System (HIS), as defined by WHO, comprises interrelated component parts for acquiring and analysing data and providing information (management information, health statistics, health literature) for the management of a health programme or system and for monitoring health activities (WHO 2000). An HIS is made up of mechanisms and procedures for providing accurate information needed by: all levels of health planners, managers for planning, programming, budgeting, monitoring, assessment and coordination of health programmes and services, health care personnel, health research workers and educators in support of their respective activities, socioeconomic planners and the general public outside the health sector for

intersectoral information linkage and national policy-makers for evidence-based policy formulation. Several phases are involved in an HIS assessment activity: assessment set-up, preparation, planning, data collection, data analysis and report preparation, follow-up and HIS plan of action preparation.

Many subsystems are included in HIS, such as a health management information system which is used to emphasise the use of the information for management of the health system, the health surveillance system and vital registration devoted to health statistics and system monitoring.

5.1.2 Health Surveillance System

The information about cancer occurrence and outcome is usually obtained from a health surveillance system (HSS), providing health information in a timely manner so that countries have the information that they need to fight epidemics in the present or plan for the future.

Surveillance is defined as the systematic measurement of health and environmental parameters, recording and transmission of data, comparison and interpretation of data in order to detect possible changes in the health and environmental status of populations, monitoring evaluates intervention, or action (Eylenbosch and Noah 1988). Disease surveillance is a continuing scrutiny of all aspects of occurrence and spread of disease to detect changes in trends or distribution in order to instigate control measures. It is considered to be a fundamental tool of public health and plays a critical role in the implementation of health care policy, both in formulating the disease control plan and in monitoring its success. The systematic collection, analysis and interpretation of information formed the basis for the description of the health dynamics of populations health and disease. This in turn is the basis for setting priorities concerning health improvement and disease control and prevention, through identifying high-risk populations or areas, and by measuring the extent and limits of a disease in a population at risk.

Three fundamental elements make up a surveillance programme: ongoing collection, analysis and feedback or dissemination of data. An effective surveillance system requires substantial and continuous effort. It should include a systematic and consistent data collecting system, a reliable and unbiased data analysis and reporting system with high validity and completeness, and also a regular, relevant and reliable feedback system. All these components should be readily assimilable, and should be up-to-date and relevant. (Eylenbosch and Noah 1988) A successful surveillance system should also offer an information service for research studies and planning exercises.

The basis for health surveillance is the acceptance of a continuing responsibility for the health situation of a population. Benefit comes only from careful analysis of the collected data, and it is therefore essential to allocate adequate resources for that purpose when a surveillance system is planned. The principal sources of health indicator data include: vital events registers, disease registers, routine health service data and epidemiological surveillance data. Sample survey methods can also be used to supplement this approach. (WHO 2005c)

A “risk factor” refers to any attribute, characteristic or exposure of an individual that increases the likelihood of developing an NCD. Emphasis in surveillance, therefore, should also be given to risk factors, especially to those amenable to intervention. Some factors not amenable to intervention, such as sex and age, are also important for estimating trends in NCDs, i.e. cancer. The WHO STEPwise approach to Surveillance of NCD Risk Factors (STEPS) is the recommended surveillance tool for measuring key risk factors for noncommunicable diseases (WHO 2003). The goal is to achieve comparability of data over time and between countries, through providing standardized materials and methods as part of technical collaboration with countries, especially those that lack resources. It also encourages the development of an increasingly comprehensive and complex surveillance system depending on local needs. In STEPS, surveillance of just seven selected risk factors (tobacco use, alcohol consumption, low fruit and vegetable intake, physical inactivity, blood pressure, cholesterol, body mass index) that reflect a large part of the future burden of noncommunicable diseases can provide a measure of the success of interventions. (WHO 2003, WHO 2005c) Three steps of surveillance are: adopting standardized questionnaires and adding modules regarding behaviour such as tobacco and alcohol use, providing physical measurements, and collecting biochemical measurements, most often by blood samples. At each step there is a core of information for each risk factor, an expanded core, and optional information, with the information of greater complexity being added sequentially as resources allow. At the country level, the implementation of this stepwise approach provides basic strategic public health information that can serve as the basis for planning and monitoring national prevention programmes as well as serving as an international standard for comparison purposes. The stepwise sequential process builds national capacity in a manner that is sustainable for the implementation of effective disease prevention programmes. (WHO 2003, WHO 2005c)

5.1.3 Cancer Surveillance System

A comprehensive national cancer control programme requires a Cancer Surveillance System providing data on a continuing basis on incidence, prevalence, mortality,

diagnostic methods, stage distribution, treatment patterns, and survival, and also the information about important risk factors and the prevalence of exposure to those factors in the population (WHO 2002a).

The major roles of cancer surveillance as indicated by WHO in 2002 are:

- to assess the current magnitude of the cancer burden and its likely future evolution
- to provide a basis for research on cancer causes and prevention
- to provide information on prevalence and trends in risk factors
- to monitor the effects of prevention, early detection/screening, treatment, and palliative care.

A surveillance system is a vital component of all cancer control programmes (WHO 2002a). It is a useful resource in epidemiological research into the etiology of disease, which is essential in the planning of prevention strategies. The establishment of an effective surveillance system requires a continuing commitment of resources, including personnel and technology, for communication, data collection and analysis.

In terms of their role in the planning and evaluation of health care, most cancer surveillance systems have proved the importance of their existence. Provision of health-care services should logically depend upon an indication of need, which can be regarded as a measure of current cancer burden under particular conditions. Projections of past cancer incidence (or mortality) trends to provide estimates of future rates of disease are also important for planning the allocation of future health resources in the local context. The use of routine cancer data to predict future need for services should, in theory at least, be a useful planning tool for national cancer control programmes.

5.1.4 Major components of a cancer surveillance system

The sources of input to a cancer surveillance system usually come from a vital statistics system, cancer registries, or an activity analysis of the health care system itself. With the pressure for cost containment in health services and variety of practice among different resources, information systems have come under scrutiny, and will increasingly need to justify their role.

Normally, information on deaths from cancer in the population is collected by civil registration systems recording vital events (births, marriages, deaths). The responsible authority varies between countries, but usually the collation of national statistics is the responsibility of the Ministry of Health of the country. Mortality data are derived from death certificates, which include information about the person dying, and in which the cause of death is certified by a medical practitioner. The International Classification of Diseases (ICD) series provide a uniform system of nomenclature and coding, and a

recommended format for the death certificate. Currently recommended by WHO is the tenth revision of the ICD. Mortality statistics are produced according to the underlying cause of death, which may not necessarily equate with the presence of a particular tumour. Approximately 72% of the world population is covered by national vital registration systems producing cancer mortality statistics. At the end of 2003, data on death registration were available from 115 countries, although they were essentially complete for only 64. (Mathers et al. 2005) This includes all of the developed countries, but few developing countries. Even when national statistics are published, their data quality varies, as indicated by e.g. completeness and coverage of registration, validity of data, consistency over time and use of ill-defined categories. (Mathers et al. 2005, Rao et al. 2005)

Several studies have investigated the validity of cause-of-death statements in vital statistics data. These studies have compared the cause of death entered on the death certificate with a reference diagnosis derived from autopsy reports (e.g. Heasman and Lipworth 1966), detailed clinical records (e.g. Puffer and Wynne-Griffith 1967), or cancer registry data (e.g. Percy et al. 1981). They have revealed that the degree of accuracy of the stated cause of death declines as the degree of precision in the diagnosis increases. Thus, although the total number of deaths from all cancer types may be only slightly underestimated, the distribution by site of cancer may be incorrect. There is a tendency to over-record non-specific diagnoses instead of the correct localization (e.g., recording the large intestine instead of the rectum as the site of cancer), and accuracy is sometimes lower in those dying at older ages or at home. There are also quite marked differences among different countries with respect to the selection of ICD code for a given death certificate diagnosis. (Percy and Dolman 1978, Percy and Muir 1989)

Cancer registration is the continuing process of systematic collection of data on the characteristics of all cancers and of the subjects with cancer. The cancer registry is an organisation for the systematic collection, storage, analysis, interpretation and reporting of data on subjects with cancer. The main objective of the cancer registry is to collect and classify information on all cancer cases in order to produce statistics on the occurrence of cancer in a defined population and to provide a framework for assessing and controlling the impact of cancer on the community. (Jensen et al. 1991) From the purely epidemiological point of view, the presence of high-risk areas for certain cancers provides good resources and opportunities for cancer research and study, for testing hypotheses about the etiology of disease and possibly for obtaining some entirely new clues. This work will be enhanced by cancer register data, too.

There are two types of cancer registry: population-based and hospital-based. A population-based cancer registry (PBCR) collects data on every subject with cancer in a defined population (usually comprising people resident in a well-defined geographical

region), designed to describe the extent and nature of the cancer burden in the population and assist in the establishment of public health priorities. (Jensen et al. 1991) The emphasis is on epidemiology and public health. The registry may be used as a source of material for etiological studies and also for helping in monitoring and assessing the effectiveness of cancer control activities. In the context of cancer burden assessment, the population-based cancer registry relates the incident cancer cases to a defined population-at-risk. The establishment of a population-based cancer registry is highly desirable in the development of a national cancer control programme. Such registries are useful in the context of documenting the cancer patterns in a given region/country, in measuring cancer burden and in studying survival from cancer as well as in evaluating trends in the incidence of cancers over time. (WHO 2002a) Hospital-based cancer registries are concerned with the recording of information on the cancer patients seen in a particular hospital. This information system provides valuable sources of information regarding methods of diagnosis, stage distribution, treatment methods, response to treatment, and survival, although accurate information on cancer incidence is unobtainable because of case referral and population coverage issues. (Jensen et al. 1991) Since these registries cannot provide measures of the occurrence of cancer in a defined population where the case originates, the hospital-based cancer registries information are not further considered in this thesis.

The main sources of information for a population-based registry include: (1) information from treatment facilities, such as cancer centres, major hospitals, private clinics, hospices, homes for the elderly and general practitioners, (2) information from diagnostic services, especially pathology laboratories, but also haematology/biochemical/immunological laboratories, X-ray and imaging clinics, (3) death certificates from the death registration system are also a very important source of information (if they are available). The information is collected from these sources in two ways: (1) Active collection, (2) Passive reporting. (Jensen et al. 1991) The cooperation of the medical profession and health care services is vital to the success of cancer registration, especially in developing countries without legislation for cancer registration.

Continuous cancer registration began in Europe in 1929, when the first population-based cancer registry was set up in Hamburg, Germany, but it was not until 1942 that the first registry to cover a whole nation was began in Denmark. This is still functioning. Since then, there has been a steady growth in the number of cancer registries and in the population covered. In 1965, the International Agency for Research on Cancer (IARC) was established as a specialized cancer research center of the WHO. One year later, the International Association of Cancer Registries (IACR) was formed, its main objective to develop and standardize collection methods across cancer registries

in the world, so as to make their data as comparable as possible. This Association collaborates closely with the IARC, compiling the ‘Cancer Incidence in Five Continents’ series every five years. As published in the latest – the 8th version, the data from 186 cancer registries in 57 countries in the world have been involved. Nevertheless, it is obvious that more registries exist in the world, although many may not reach the data quality requirement of the CI5 series, or are not represented in the above organisations.

5.2 Sources of information about cancer statistics in the world

Estimation or prediction of global cancer statistics is one of the tasks of IARC/WHO, where scientists have used the latest available cancer information from different countries in the world to present statistics describing the global cancer burden. Updated global and regional cancer statistics have been continuously released in various scientific publications, and also in software packages. Nevertheless, Parkin et al. (2001) indicated that it was natural that the accuracy of the estimates varied widely between the world areas, depending on the nature and extent of the information on cancer occurrence or outcome available in the component countries. There was also considerable heterogeneity of cancer rates between countries within the same area or within the larger countries.

5.2.1 The CancerMondial Project

The Cancer *Mondial* Project was developed by the Descriptive Epidemiology Group in IARC/WHO and described on its website (IARC 2005). It is a joint initiative of IARC and the European Union (EU). The aim of this project is provide online access to data on the incidence, prevalence, survival and mortality of cancer held by IARC. The strategy is to gradually increase the scope of the information available, while developing an interface which will allow the user to make highly specified queries based on combinations of populations (countries or regions), sex, age, types of cancer and outcome (e.g. numbers of new cases, prevalence or mortality rates). Within this project, there are major information resources about cancer statistics or burden published on the world-wide, regional, country or registry levels with different formats: electronic database, software, technical reports, or scientific publications (books and papers). Estimates of the worldwide incidence of and mortality from major cancers have

been prepared every 5 years since 1975. There is also an international incidence of childhood cancer series, cancer occurrence and survival in developing countries, cancer in Africa and some cancer statistics at the local country level such as in Portugal, Spain, Thailand, France, Zimbabwe, Philippines and Costa Rica. (<http://www-dep.iarc.fr/publi/publica.htm>).

(1) **Cancer Incidence in Five Continents series:** The aim of the Cancer Incidence in Five Continents (CI5) series is to describe cancer incidence in a comparable fashion for all the populations in the world for which good quality data are available. A volume in the series of monographs has been published every five years. Cancer incidence data in the CI5 series are derived from cancer registries around the world that record the occurrence of cancer over a specified period, following common guidelines about cancer registration methods. The data are subjected to quality checks for comparability, completeness and validity (Parkin et al. 2002, IARC 2005) before inclusion, and many submitted datasets fail to meet the standards. The CI5 series is widely used as a data source on international incidence of cancer. CI5 is the result of the collaboration between the IARC and the network of cancer registries worldwide, represented by the International Association of Cancer Registries (IACR). So far, eight volumes have been published, covering a 40-year period, and providing information on cancer patterns geographically, by ethnic group and over time (Parkin et al. 2005b). There has been an improvement in the availability of high quality cancer incidence data in the world during last 50 years. This is reflected in the CI5 series, which has grown from the 32 cancer registries from 29 countries represented in the first volume to 186 registries in 57 countries included in the latest volume. The authors have drawn attention to the potential for understanding disease etiology that can be derived from the study of disease frequency in different areas and over time. (Doll et al. 1966, Doll et al. 1970, Waterhouse et al. 1976, Waterhouse et al. 1982, Muir et al. 1987, Parkin et al. 1992, Parkin et al. 1997, Parkin et al. 2002)

(2) **GLOBOCAN database: Cancer Incidence, Mortality and Prevalence Worldwide:** The GLOBOCAN database is based on the most recently available data on cancer mortality, incidence and survival worldwide. Two versions have been developed: GLOBOCAN 2000 and a newly updated version – GLOBOCAN 2002, which is now available on the IARC website: <http://www-dep.iarc.fr/dataava/globocan>. GLOBOCAN 2002 presents estimates of the incidence, prevalence of and mortality from 27 cancers by sex and age for year 2002 from all countries in the world. Different methods to estimate the sex-, age-, and site-specific rates of cancer have been used according to the extent and accuracy of locally available data in different countries, such as national mortality statistics and

cancer registries. The methods are fully described in the GLOBOCAN 2002 package (Ferlay et al. 2004) and are summarized by Parkin et al. (2005a). To estimate incidence, for example, national cancer registry data are the preferred source, but, in their absence, mortality data are used, and incidence is estimated by applying site-specific regression models of incidence as a function of mortality, obtained from Poisson regression analyses of incidence and mortality data provided by cancer registries in the same country or region, with adjustment for age and sex (Bray et al. 2002). It should be noted that the mortality data from the registries must be from the same source(s) as the national mortality data. National incidence data can be converted into mortality rates using data on survival from the same country or region. In the worst case, when no data are available on cancer incidence or mortality (for some of the developing countries in Africa), a set of age/sex specific incidence rates for all cancers (based on rates observed in countries in the same region) is partitioned between different cancer sites using any available data on the relative frequency of different cancers (by age and sex). The numbers of cases, deaths and cancer survivors are computed by multiplying the estimated rates by the population estimates for the corresponding country by the United Nations. Due to the variable delays in collection of cancer data, compiling and reporting, the estimated rates are not those for the year 2002, but from the most recent period available, generally 2–5 years earlier. Due consideration should be given to the variable degree of detail and quality of the data among the countries when interpreting the estimated rates. (Parkin et al. 2001, Ferlay et al. 2001, Ferlay et al. 2004, Parkin et al. 2005a)

5.2.2 Other cancer statistics in regions, areas or countries

In 1981, Michael Alderson published a book, giving information about statistics on mortality in the 20th century for European and other selected countries (total of 31 countries involved). He addresses the importance of considering the validity of the data when handling mortality statistics and before attempting to interpret any trends in disease. Comprehensive consideration of statistical methods, and the indicators used for calculating mortality (both population estimates and the cause of death for individuals), as well as various factors may influence the validity of data and interpretation of statistics are discussed in detail in the book. Various indicators of the validity of mortality statistics were stressed in that book: the proportion of doctors in a population, the proportion of the population living in rural areas, the percentage of death certificates issued by medical practitioners and the percentage of all certificates that are from ill-defined causes of death. (Alderson 1981) It describes the uses of mortality statistics in

descriptive studies, generation of hypothesis, hypothesis testing and evaluation of public health measures. It also points out that the mortality statistics should facilitate specific studies of the impact upon mortality of national programmes to control particular diseases.

The national cancer mortality data used in GLOBOCAN were derived from the WHO Mortality Database (WHO 2005b), extracted from the WHO databank through the WHO Statistical Information System (WHOSIS) (WHO 2005b). This mortality database contains mortality data officially reported by WHO member states and includes cause-of-death statistics coded according to the 9th and 10th revision of the ICD. The data represent official national statistics that have been transmitted to the WHO by the authorities of the countries concerned. The 2002 revision of world population prepared by the UN Population Division (UN-PD 2003) is currently used for national population estimates. Death rates by age and sex are applied to the national population data to obtain total estimated deaths. Nevertheless, some information on the likely accuracy of the data is a prerequisite to their appropriate use. The publication by national and international authorities is not a guarantee of data quality. In some countries or time periods, population coverage is manifestly incomplete, and the reported mortality rates produced are implausibly low. The WHO Statistical Information System publishes tables of estimated coverage and completeness of mortality statistics in their database. Due to these deficiencies, quality of cause of death information may be poor. This is obvious when a substantial proportion of certificates is completed by non-medical staff. WHO formerly published quality indices for some countries in the World Health Statistics Annual (WHO 1998). Otherwise, quality of data must be judged from indicators such as the proportion of deaths coded to “senility and ill-defined conditions,” and the proportion of cancer deaths without specification of the primary site, or for which site is specified in only vague terms. Mathers et al. (2005) recently reviewed the availability and quality of worldwide data on death registration, with respect to data timeliness, completeness and coverage of registration, the coding system used for causes of death, and the proportion of deaths coded to various ill-defined categories. Relatively few countries were found to have good-quality data on mortality that can be used to adequately support policy development and implementation (Mathers et al. 2005).

A five-year Global Burden of Disease Study (GBD) was initiated in 1992 by the World Bank, in collaboration with the WHO, aimed at providing estimates of non-fatal health outcomes, to give unbiased epidemiological assessments for major disorders, and to quantify the burden of disease with a measure that could also be used for cost-effectiveness analysis. Two GBD studies have been conducted so far: in 1990 and 2000. In the GBD 1990 study, the basic units of analysis were the eight World Bank Regions

defined for the 1993 World Development Report (World Bank 1993). After a literature review, disease experts developed first-round estimates, which could then be critically reviewed. The internal consistency was ascertained with a computer program (DisMod). Major inconsistencies were identified and estimates revised to correct them. Three groups of causes of mortality were classified in the study – communicable diseases, non-communicable diseases and all injuries. The primary indicator used to summarize the burden of premature mortality and disability was the disability-adjusted life year (DALY). The burden of 107 disorders was compared with the burden attributable to ten major risk factors and to selected diseases as risk factors for other conditions. One major finding from this study was the importance of non-communicable diseases in all regions of the world. (Murray et al. 1994, Murray and Lopez 1996, Murray and Lopez 1997, Lopez and Murray 1998) In the GBD 2000 study, health state valuations were determined (population-based data rather than expert opinion as used in the 1990 study) and an approach for modelling the relationship between the level of mortality and the disease structure in populations, based on proportions rather than rates, was used. The GBD 2000 project classified WHO's 6 regions into 17 sub-regions according to the levels of child and adult mortality. Estimates of disease and injury burden were first developed for each individual Member State of WHO, using different methods for countries at different stages of health development, often largely determined by the availability of data. The regional site-specific cancer mortality was obtained by disaggregating the regional cancer mortality for all cancers in the region, based on the mortality distribution by site estimated from vital records or cancer survival models. For regions with high coverage and ICD-coding vital registration system, such as most countries in Europe, America, Australia, Japan and New Zealand, direct estimates of the site-specific distributions of cancer mortality were used from their registration. For other regions of the world, a site-specific model for relative interval survival adjusted for each region was developed and applied to the regional incidence from GLOBOCAN 2000 to calculate the site-specific mortality distribution in 2000. The National Cancer Institute Surveillance, Epidemiology, and End Results (SEER) statistical program (SEER*Stat version 4.2) was used for developing the cancer survival model in the GBD 2000 study. (Mathers et al. 2002, Shibuya et al. 2002, Lopez 2005, WHO 2005e)

In Europe, the European Network of Cancer Registries (ENCR) was launched in 1989 within the framework of the 'Europe Against Cancer' programme of the European Community. The objectives of this Network is to improve the quality, comparability and availability of cancer incidence data, create a solid basis for continuous monitoring of the cancer incidence and mortality in the European Union, provide regular information on the burden of cancer in Europe and to promote the use of cancer registries in cancer control, health-care planning and research. (ENCR 2005) The European Cancer

Incidence and Mortality Database (EUROCIM) is shared among only eighty-four cancer registries within the network, describing the cancer incidence, mortality statistics, covering approximately 40% of the population of the Community. Complete information is available for some countries but, for others, some form of estimation is required. (ENCR 1995) The latest version of EUROCIM database includes a time trends analysis module. This new feature allows the user to fit age-period-cohort models to the registry incidence and mortality data. (ENCR 2001) Various publications estimating cancer incidence and mortality in different populations have been released in recent decades (Estève et al. 1993, Black et al. 1997, Ménégos et al. 1997, Miñarro et al. 2000, Bray et al. 2002, Pinheiro et al. 2002). There is an additional collaborative project shared out with the Network – CaMon project (Comprehensive Cancer Monitoring in Europe) extending the role of the ENCR (The European Commission Health Promoting Programme 2005). It aims to develop a cancer surveillance system for cancer occurrence and outcome, permitting situation analysis and monitoring of cancer burden in the Member States of the European Union and applicant states, to disseminate such information within the European Union and world-wide, and to make it available for incorporation into the health monitoring system of the European Union Public Health Programme. Three major cancer databases present the cancer statistics in Europe generated by this project: EUCAN, EUROPE 95 and EUROCARE.

Besides the cancer statistics shown on the ENCR website, various resources for cancer information exist for individual countries or regional groupings. A good example is the Association of the Nordic Cancer Registries (ANCR), which is the cooperative organisation for the national cancer registries of Denmark, Finland, Iceland, Norway and Sweden. In cooperation with IARC/WHO, this Association developed NORDCAN software, a unique source of the information on cancer incidence and mortality in the Nordic countries, using incidence and mortality data provided by cancer registries in the Nordic countries. (Storm et al. 2003) NORDCAN includes data on 41 major cancer sites that can be presented as a variety of tables and graphs, and that can be easily exported or printed. NORDCAN allows regions and cancer sites to be grouped and compared as desired. Furthermore, each cancer registry has its own website to show the cancer statistics in its country.

In North America, the North American Association of Central Cancer Registries (NAACCR) (2005) was established in 1987. It is a collaborative umbrella organisation for cancer registries, governmental agencies, professional associations, and private groups in North America interested in enhancing the quality and use of cancer registry data. All central cancer registries in the United States and Canada are members. Cancer reporting is mandated through state or provincial laws. Registries from each state or province voluntarily submit their non-identified data for evaluation and publication to

NAACCR. After NAACCR receives the data, Cancer in North America (CINA), a monograph containing most publications about cancer incidence rates and supporting data from participating states and provinces in USA and Canada, is published, and an interactive query system called CINA+ Online has also been developed – NAACCR STATISTICS & REPORTS. (NAACCR 2005)

Cancer statistics in the United States can be found from several different sources. The authoritative source of information on cancer incidence and survival in US is the Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute (NCI) which routinely collects data on patient demographics, primary tumour site, morphology, stage at diagnosis, first course of treatment, and follow-up for vital status (NCI-USA 2005). SEER began collecting data on cancer cases from 1973 in 5 States and 2 metropolitan areas, and currently collects and publishes cancer incidence and survival data from 14 population-based cancer registries and three supplemental registries covering approximately 26 percent of the US population. Information on more than 3 million *in situ* and invasive cancer cases is included in the SEER database, and approximately 170,000 new cases are added each year. Computer applications have been developed to unify cancer registration systems and to analyse and disseminate population-based data in the US and also in the world. The Annual Cancer Statistics Review (CSR) reports the most recent cancer incidence, mortality, survival, and prevalence statistics from the SEER registries. The most recent report is SEER CSR 1975–2002, which includes incidence, mortality, survival, and prevalence statistics from 1975 through 2002, and can be download from the web – http://seer.cancer.gov/csr/1975_2002 (Ries et al. 2005). Cancer mortality information for the United States and for individual states is based on causes of death reported by the certifying physicians on death certificates filed in state vital statistics offices. The mortality information is processed and consolidated into a national database – Mortality Data through the National Vital Statistics System (NVSS) developed by the National Center for Health Statistics (NCHS) in the Centers for Disease Control and Prevention in US (CDC) (CDC-USA 2005). NVSS results from an inter-governmental collaboration between NCHS and the 50 states, two cities, and five territories where vital events – births, deaths, marriages, divorces, and fetal deaths are registered by the legal authority. Shared relationships, standards, and procedures between these organisations form the mechanism by which NCHS collects and disseminates mortality statistical information about the nation's official vital statistics from death certificates. These jurisdictions are responsible for maintaining registries of vital events and for issuing copies of birth, marriage, divorce and death certificates. US cancer mortality can be directly analysed with the NCI's SEER*Stat software. (NCHS-USA 2005) Besides NCI/NIH, an annually published Cancer Facts & Figures by the American Cancer

Society (ACS)'s Epidemiology and Surveillance Research Department are also available online, in which data on cancer incidence, mortality, survival, and also the risk factors are presented and the numbers of new cases and deaths estimated (ACS 2005). All these cancer statistics estimates, as Greenlee et al. (2000) indicated, due to the delays inherent in the release of incidence and mortality data and no comprehensive national cancer registry available in the United States, provides a useful way to evaluate the current burden of cancer on the US and individual state populations. But caution should be exercised when interpreting the estimates of the expected numbers of new cancer cases and cancer deaths tracking trends over time since the estimates may vary considerably from year to year and may not really represent the current burden of cancer, data normally based on information several years ago. Cancer incidence statistics in the US are usually based on incidence rates for the geographic location that participate in the SEER program and therefore, may not be entirely representative of the US as a whole.

The major source of cancer statistics in Australia is the National Cancer Statistics Clearing House (NCSCCH) at the Australian Institute of Health and Welfare, a national agency for health and welfare statistics and information which was established in 1987 (NCSCCH 2005). The main objective of NCSCCH is to foster the development and dissemination of national cancer statistics for Australia and specifically to enable computation and publication of national statistics on cancer and facilitate cancer research both nationally and internationally. The NCSCCH receives data from individual State and Territory cancer registries on cancer diagnosed in residents of Australia, which enable a national dataset to the established following record linkage of cancer by site and behaviour. This commenced with cases first diagnosed in 1982. NCSCCH produces reports of national incidence and mortality and periodic analyses of specific cancer sites, cancer histology, differentials in cancer rates by country of birth, geographical variation, trends over time and survival. Information about cancer incidence data for Australia for 1983–2000 has been released on the interactive cancer data page - <http://www.aihw.gov.au/cancer/ncsch/index.cfm>. On a state basis, a range of statistics is available from state and territory cancer councils, including the Cancer Council Victoria and the Cancer Council Australia. Cancer mortality data can be extracted through the National Death Index database in the Australian Institute of Health and Welfare, which contains records of all deaths occurring in Australia since 1980.

The official cancer statistics information in Japan – "Cancer Statistics in Japan" is published biennially by the Foundation for Promotion of Cancer Research. Data for the period 1997–2003 can be found through the website of the National Cancer Center in Japan by utilising the results of the two comprehensive 10-Year Strategy for Cancer

Control during 1984 and 2003. (NCC-Japan 2004) In South Korea, Cancer Statistics in Korea is published by the Cancer Registration & Biostatistics Branch in National Cancer Center in Korea. The data include the cancer mortality statistics based on annual reports from vital registration, and incidence information from both central and regional cancer registries. (NCC-Korea 2005)

In most other countries in Asia, cancer incidence information is available from fewer cancer registries' usually covering only a fraction of the national population, and often with lower quality or less information than data coming from the developed countries. Nevertheless, as shown in Cancer *Mondial* Projects in IARC, with the help of IARC, the cancer statistics in some countries have been published in various scientific papers or technique reports, mainly based on the regional cancer registry data, such as in Thailand, India, Vietnam and Pakistan (Quoc et al. 1998, Deerasamee et al. 1999, Bhurgri et al. 2000, Sen et al. 2002). The sources of cancer statistics in China are described separately under Section 5.3.

Knowledge of cancer pattern in Africa is family limited due to the lack of valid mortality statistics, wide diversity of genetic population, social-economic status and environmental factors. Cancer in Africa compiled by Parkin et al. (2003), includes a description of all cancer registration activity ongoing in Africa today, as well as in the past. All of the currently available data on the profile of cancer and a review of the published literature are provided for every country (52 countries in total) in Africa, notably from cancer registries where data have not hitherto been published or have appeared only as journal articles or in local reports. Furthermore, there is a comprehensive review of the epidemiology and prevention of 20 major cancers of importance on the African continent, together with chapters on cancer in children, and cancers related to AIDS. (Parkin et al. 2003)

5.3 Sources of information on cancer statistics in China

China has a long history of medical care, and data about cancer incidence and mortality have been collected for several decades. These data, however, has mainly been published in Chinese, and often covered a rather small population (Tu 1985, Wang et al. 1995, Gao 1991, Chen 2000). The vital information (birth and death, with legal provision for registration) is collected by a national civil registration system – the National Household Registration System established in 1954 and run by the Public Security Department. However, currently a qualified report on the cause of death is available for only about one tenth of the Chinese population (Yang et al. 2005). The National Office for Cancer Prevention and Control (NOCCPC) was established in Beijing

in 1969, aiming to promote cancer prevention and control in the whole country through support from both technical and administrative aspects. In recent decades, the NOCPC has made major efforts to promote the development of a cancer information system in China, both for cancer mortality and incidence registration.

The sources of information on cancer in China fall into three broad groups: (i) mortality data from national retrospective surveys, (ii) mortality data obtained through routine reporting systems or special research projects on the causes of death, and (iii) incidence and mortality data from cancer registries (Yang et al. 2003b). These data have been used, at different levels, to estimate the pattern or burden of cancer in China (World Bank 1993, WHO 1997, Parkin et al. 1999, Pisani et al. 1999, Ferlay et al. 2001, Pisani et al. 2002, Ferlay et al. 2004).

5.3.1 Mortality data from two national retrospective death surveys

After the establishment of the NOCPC, two national retrospective surveys of the cause of death were organised and conducted by this office, providing a clear picture of the distribution of mortality rates for the principal diseases affecting the Chinese people and a sound basis for establishing priorities for research in basic and clinical sciences both for Chinese and global cancer researchers.

The first national mortality survey was conducted in 1976, and covered about 850 million people (96.7% of the Chinese population). It involved identification of about 20 million deaths during the years 1973–1975, retrospective diagnosis of their causes. It provided for the first time an indication of the sex- and age-specific mortality rates for the principal diseases affecting the Chinese population. Using a uniform questionnaire, nearly one million health workers obtained information from various levels of the medical system. The accuracy of data varied between regions and source of information. Checking and rechecking based on samples was carried out to guarantee data completeness and reliability. The demographic information was provided through the vital statistical system, while a uniform questionnaire for cause of death was used to obtain the death information from different levels of the medical system (village, commune, county, district in prefecture/city, and province). The causes of death were classified into 20 categories for 56 diseases with 15 subgroups malignant tumours: nasopharynx, oesophagus, stomach, liver, lung, breast, cervix uteri, leukemia, colon/rectum, bladder, penis, lymphoma, choriocarcinoma, brain tumour and other malignant tumour. (NOCPC 1979) Of disease deaths, 70.5% had an ante mortem diagnosis based on either pathology, cytology, bone marrow examination or a clinical investigation (including exploratory laparotomy without biopsy, x-ray, ultrasound, isotope scans, and biochemical and immunological laboratory tests). During those three

years, cancer was the second leading cause of death, following respiratory disease for males, and in third place for females, after respiratory, heart diseases other than coronary heart disease and cerebrovascular disease. The five most common cancers in men were cancers of the stomach, esophagus, liver, lung and colon/rectum, while in women they were stomach, cervix uteri, oesophagus, liver and lung. The geographic distribution of mortality for common cancers was published in an atlas (Editorial Committee for the Atlas of Cancer Mortality 1979). According to Doll and Greenwald – “Few medical projects can ever have been successfully carried through that compare in scale with the national survey of the causes of death undertaken in China in 1976” (Chen et al. 1990). A map of the distribution of mortality from nasopharyngeal cancer in 1973–1975 is shown in Fig. 1 (Editorial Committee for the Atlas of Cancer Mortality 1979).

The second retrospective mortality survey was conducted in 1993, and comprised the causes of death during the period 1990 to 1992. Instead of a nationwide survey, a 10% sample survey was conducted, using a two-stage randomized stratified cluster sampling method. The first stage sampling was done on provincial/municipality level, then the second stage on the county/town level. In each stage, three strata (high,

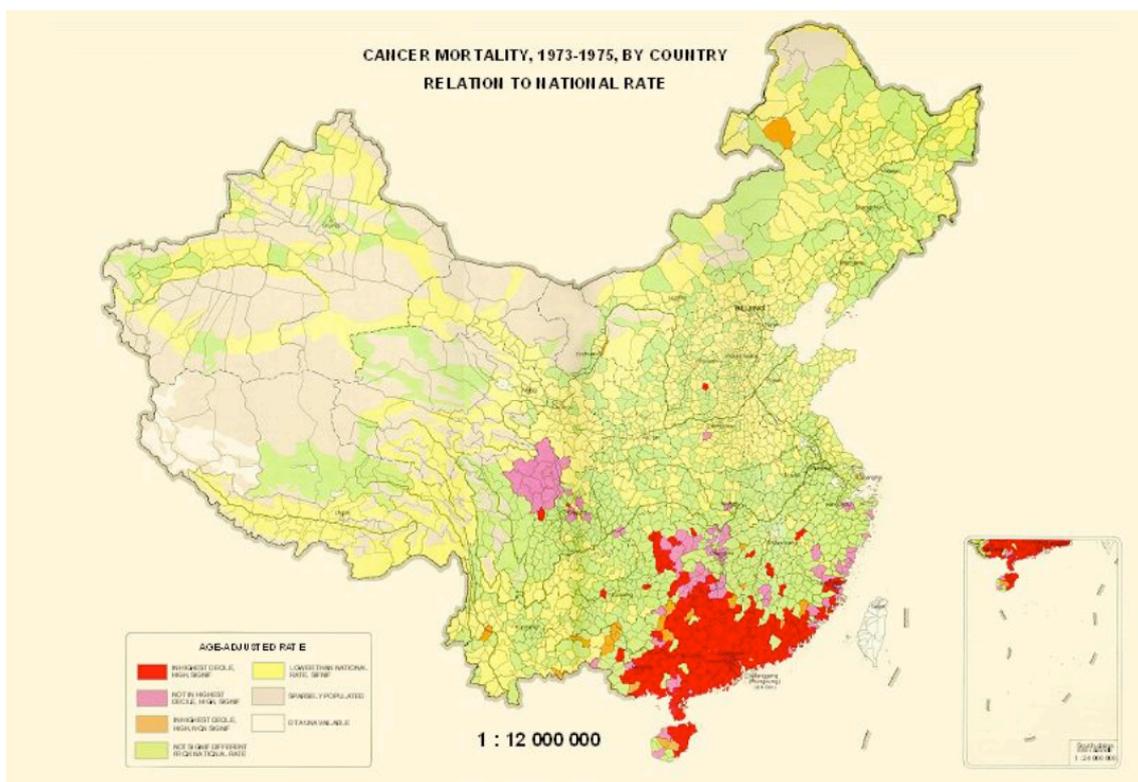


Fig 1. Map of nasopharynx cancer mortality in men in China, 1973–75, from the Atlas of Cancer Mortality in the People’s Republic of China, 1979

medium and low mortality) were defined based on the mortality rates in the survey of 1973–1975. A total of 263 sample units (74 cities and 189 counties) among 27 provinces finally was selected, covering 10% of the total Chinese population. Quality control included training programmes at different levels before the survey, data checking according to the indicators listed in a questionnaire related to the data completeness, validity and reliability developed by NOCPC, throughout the whole survey. Further 2% sampling re-checking was based on the household re-interview after the data collection in each sample unit (county or town). The data completeness, representativeness and reliability have been found acceptable (Li et al. 1996). Of the deaths, 31.8% had a diagnosis based on pathology evaluation, and about 60% of others had a cause of death based on autopsy, surgery, clinical investigations, and clinical chemistry estimation. Cancer still ranked second among all causes of death in this survey, after respiratory disease. The geographic distribution of cancers in China was similar to that observed in the 1973–1975 survey. Based on the results of this survey, the cancer mortality in the early 1990s and changes in the cancer mortality pattern since the first survey in the 1970s have been analysed in several publications. (Li et al. 1996, Li et al. 1997a, Li et al. 1997b, Zhou et al. 1997, Sun et al. 2002, Chen et al. 2003, Yang et al. 2003a, Sun et al. 2004)

The differences in the age standardized mortality rates (using the Chinese census data in 1982 as the standard population) for eight common cancers between the two surveys are shown in Fig. 2. The mortality rates for all cancers increased by 11.6% (from 84.6 per 100,000 to 94.4 per 100,000). With respect to the different cancer sites, the most common cancers increased during the two decades, especially lung cancer (from 7.2 per 100,000 to 15.2 per 100,000, an increase of 90%, for both sexes), while a dramatic decrease was observed for cervical cancer (from 10.3 per 100,000 to 3.3 per 100,000, decreased by 68%).

The age sex-specific mortality rates for all cancers combined in the two surveys are shown in Fig. 3, on a logarithmic scale. Mortality in middle-aged women was lower in 1990–1992 than in 1973–1975, probably reflecting the dramatic decline in mortality from cervical cancer. But the most obvious change is the increase in mortality rates in the older age groups (after age 55 in men and 60 in women).

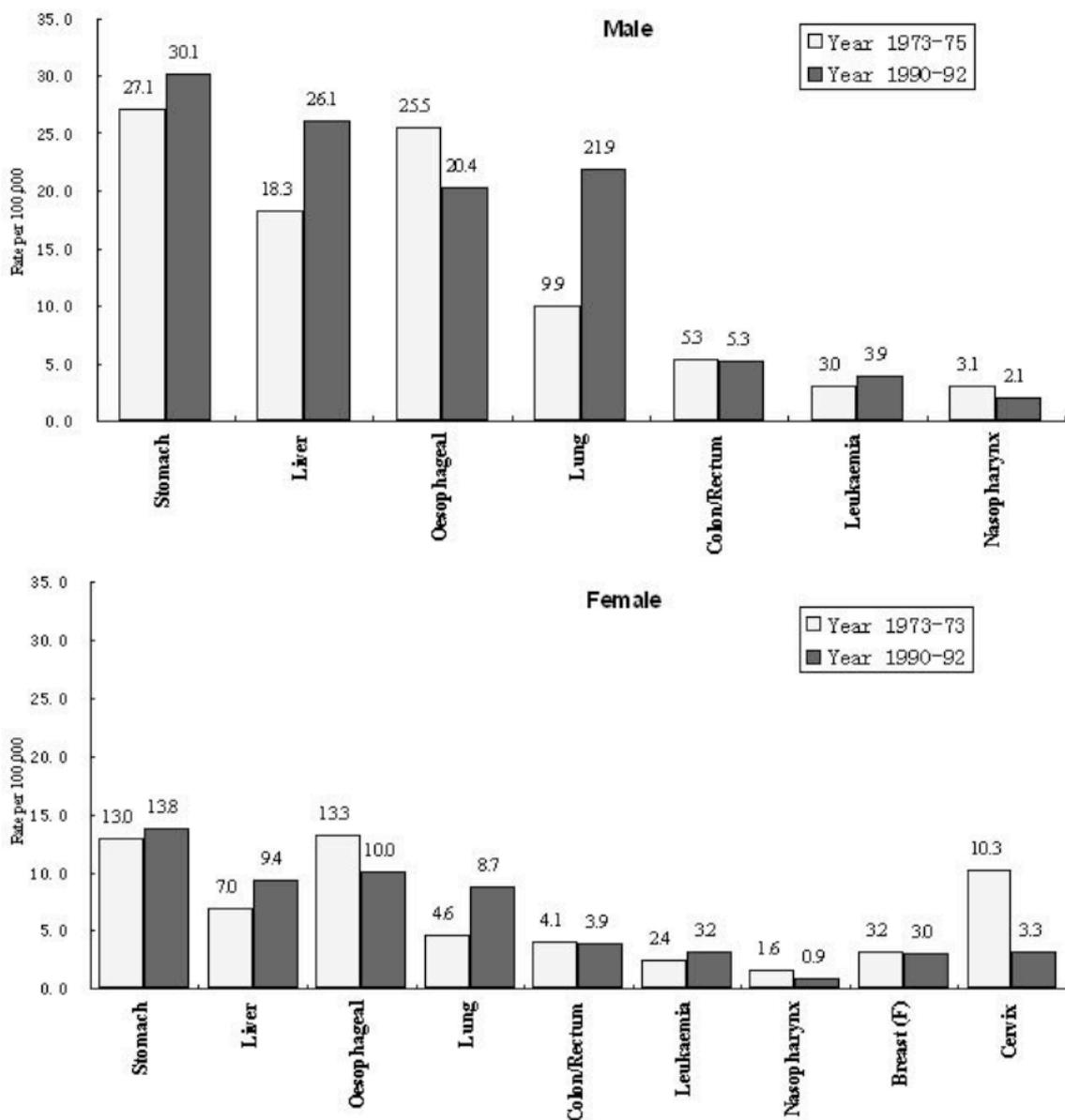


Fig.2 Differences of the age-standardized mortality rates for cancer in year 1973–75 and year 1990–92, by sex, in China

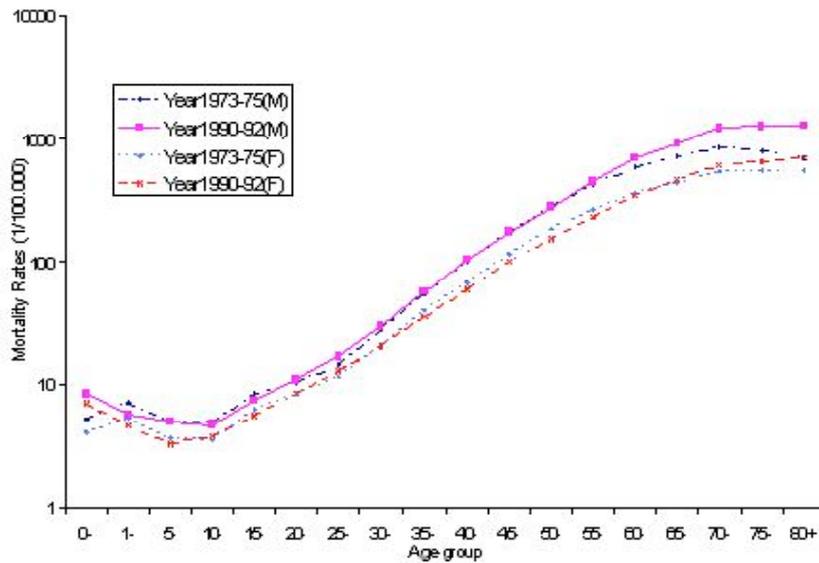


Fig. 3 Age-sex-specific Mortality Rates for Cancer in China (two national mortality surveys in 1973–75 and 1990–92)

Nevertheless, the cancer mortality profiles among urban and rural populations are quite different (Fig. 4). The mortality from lung cancer was much higher in urban than in rural areas for both sexes. In 1990–1992, it was the most common cause of death from cancer in urban areas. In rural areas, cancers in the stomach, liver and oesophagus were the most common cause of cancer death for both sexes. In addition there was very high mortality from cervical cancer in the 1970s. During the two decades, the age-standardized mortality rate (adjusted by the Chinese population census in 1982) for all cancers combined remained stable in urban areas (90.9 per 100,000 in the 1970s and 89.9 per 100,000 in the 1990s). However, in rural areas, the rates were considerably higher in the later period (81.7 per 100,000 in 1970s vs. 96.5 per 100,000 in 1990s). As for mortality changes for specific cancer sites, dramatic increases were seen for cancers in lung and liver, as well as leukemia for both sexes in both urban and rural areas, while changes were more remarkable in rural than in urban areas. Decreases were found for cancer in nasopharynx for both sexes and in cancer of the cervix in women (sharper changes for urban women). Better diagnostic facilities and changes in coding practices contributed to an increasing proportion of cancers of the gastro-oesophageal junction being coded to stomach (cardia), rather than oesophagus (as in the 1970s), which makes interpretation of the differences in mortality rates for these two cancers between the series problematic. However, declines were observed in mortality rates from both of these cancer types in urban areas, for both sexes. Changes in the socio-economic and health situation, life-style, and environmental factors between urban and rural areas during those two decades are most likely responsible for the differences observed between these two surveys. (Li et al. 1997a)

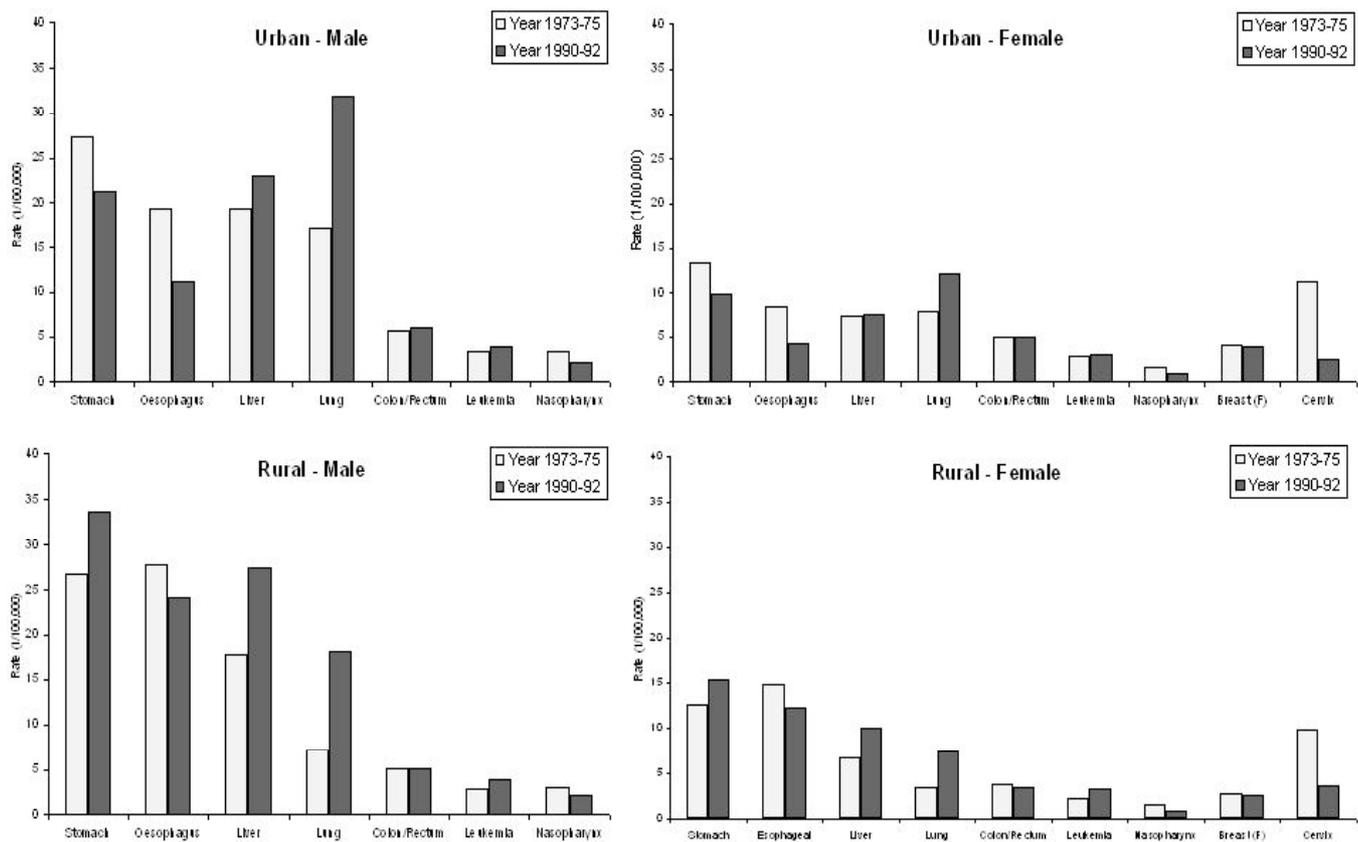


Fig. 4 Age-standardized mortality rates for cancer 1973–75 and 1990–92 in urban and rural areas by sex in China

5.3.2 Mortality data from routine reporting systems or special projects

There are two large ongoing routine reporting systems for the cause of death in China (Yang et al. 2003). One is the mortality data from the Center of Health Information and Statistics (CHIS) in the Ministry of Health, which was set up in 1973 and is based on a 10% sample of the population (between 100 and 120 million persons). The CHIS sample population comprises 15 big cities, 21 middle-sized or small cities and 85 counties (WHO 1994, NCHIS 1999, SSBC 1999, CHIS-MOH 2002, Yang et al. 2003). Many geographical areas are voluntarily participating in the data collection exercise, resulting in many sample sites located on the eastern seaboard, where population density is high (Rao et al. 1992, WHO 1994, NOCPC and NCHIS 2001).

The reporting procedure for mortality information is as follows: After a person dies, the family member reports the death to the registration office to obtain a death certificate, which is a document necessary for deregistering a permanent resident in the police station, and to obtain a permit for burial. Based on the description from a family member and available medical records or documents, the death certificate is filled out

by the staff of the local vital registration office. One copy of the certificate is kept in the registration office, and another is sent to the county Center for Disease Control (CDC), where the cause of death is coded. The county CDC then submit monthly reports of deaths, in a prescribed summary tabulation format, by age, sex and cause to the Center of Health Information and Statistics. (Yang et al. 2005) A Chinese Classification of Diseases List consisting of more than 500 diseases or injuries was formerly used at the beginning in the 1980s. However, it can be converted into the ICD coding system. Since 1990, ICD-9 has been used. Soon after 2000, the MOH required that ICD-10 be used for coding a disease in the country.

CHIS data are submitted to the WHO Mortality Database, in which cancer mortality from years 1987 to 1999 are available (WHO 2005b). The sample population is not representative of the national population with respect to socio-economic status. Rao et al. (1989) divided health status in China into 4 classes; the 85 sample counties included no counties from the lowest class.

The other routine mortality reporting system, Disease Surveillance Points (DSP) has been run by the Chinese Academy of Preventive Medicine (CAPM, currently the Chinese Center of Disease Control) since 1980. It includes 145 surveillance points at present, almost 1% of the population in China. This system was initially intended for reporting acute infectious disease. Most sites were located in the big cities, or in counties with high socio-economic status (the first stage). In 1989, a second stage of the national DSP was introduced, using a multistage stratified sampling method in an attempt to obtain a more representative sample at the national level (NOCP and NCHIS 2001, Yang 1992). Geographic, administrative and various health indicators (GNP, literacy rate, birth rate, infant death rate, crude death rate, proportion of the population in age-group 0–14, proportion of the population over 65) were used as factors when selecting the sampling population. The information on mortality is based on the death certificates. An estimate was made for testing the degree of under-reporting, which was used to adjust the observed mortality rates. (Yang 1992, Yang et al. 1996, DCD-MOH and CAPM 1999) The mortality data from the DSP system have been used in a World Bank report (World Bank 1993).

Besides the two routine mortality reporting systems described above, some extensive research projects have provided useful cancer mortality information on the Chinese population, for example, the two national tobacco hazard studies in China around 1989–1991 (one is retrospective, the other is prospective), to monitor the evolving epidemic of mortality from tobacco in China, conducted by CAPM, the Chinese Academy of Medical Sciences (CAMS) and the Clinical Trial Service Unit (CTSU) in UK (Liu et al. 1998, Niu et al. 1998). The retrospective study was carried out during 1989 and 1991 in 98 areas in China (24 cities and 74 rural counties). The cause

of death and smoking habits of one million people who died during 1986–1988 was recorded by interviewing the surviving family members (or sometimes, in rural areas, other informants). In the prospective study, the mortality information was monitored through local residential records and the causes of death recorded from the death certificates.

5.3.3 Incidence and mortality data from cancer registries

The development of cancer registration in China has been slow and laborious, although there has been great progress during last decade. In 1963, the first cancer registry in China was established in Shanghai. From the 1970s, more registries were set up with the support of NOCPC. In 1995, a ‘Chinese cancer incidence, mortality and risk factor surveillance programme’ was conducted in eleven cities/counties by NOCPC and CHIS, including 5 big cities (Beijing, Tianjin, Shanghai, Wuhan and Haerbin) and 6 high-risk areas for certain cancers (Cixian, Linzhou, Changle, Qidong, Jiashan, Fusui). This programme was approved, and dramatically improved cancer registration work in the areas. (Li et al. 2000) Three monographs have been published (NOCPC and NCHIS 2001, NOCPC and NCHIS 2002, NCHIS and NOCPC 2003) which provide a good reference for other cancer registries’ practice. In the 8th edition of Cancer Incidence in Five Continents (Parkin et al. 2002), data from eight cancer registries involved in that programme (Beijing, Shanghai, Tianjin, Wuhan, Qidong County, Jiashan County, Changle County, Cixian County) were published. The cancer registry data from China that have been published in successive volumes of CI-5 series are shown in Table 1 (Waterhouse et al. 1982, Muir et al. 1987, Parkin et al. 1992, Parkin et al. 1997). The distribution of cancer registries where data published in the 8th version of CI-5 is shown on the map in Fig. 5. The ‘Cancer Registry and Surveillance Association’ and the ‘National Center of Cancer Registries’ (NCCR) were established separately in 2001 and 2003, to set up a national cancer registry network. Referring to the principle and methods of cancer registration provided by IACR, the experience in other countries, combined with the local context in China, based on a probative norm for cancer registration developed in 1988 (NOCPC 1988), a ‘Guideline for Chinese Cancer Registration’ was published by NOCPC, NCHIS and NCCR (2004), as an official handbook for the cancer registries in China. However, it is known that there are many registries established by local institutions in China that are not included in the NOCPC’s registry network, and therefore receive no central technical or financial support, and the quality of the data collected has not been evaluated. (NOCPC and NCHIS 2001, Yang et al. 2003b)

Despite all the sources of information on cancer incidence and mortality for subsets of the Chinese population, and the two national mortality surveys, little previous research effort has been devoted to systematically comparing the quality of these information resources before our study. However, one very recent study (Rao et al. 2005) has done so, based on a criteria-based evaluation for the quality of both CHIS and DSP datasets. The criteria used in that study were related to the data generalizability (coverage and completeness), reliability, validity and also the policy relevance of the data. It was concluded that mortality registration is incomplete in both (CHIS and DSP datasets). No statistics were available for geographical subdivisions of the country to inform resource allocation or for the monitoring of health programmes, although the study did find that both datasets showed consistency of mortality data with time. (Rao et al. 2005)

Table 1. Cancer registries included in different volumes of Cancer Incidence in Five Continents series (Waterhouse et al. 1982, Muir et al. 1987, Parkin et al. 1992, Parkin et al. 1997).

Cancer Registry	Cancer Incidence in Five Continents				
	Vol. IV	Vol. V	Vol. VI	Vol. VII	Vol. VIII
Shanghai	1975	1978-82	1983-87	1988-92	1993-97
Tianjin		1981	1983-87	1988-92	1993-97
Qidong			1983-87	1988-92	1993-97
Beijing					1993-97
Jiashan					1993-97
Wuhan					1993-97
Changle					1993-97
Cixian					1993-97

←→ : for all cancer sites; ←---→ : for certain cancer sites only



Fig. 5 Map of Chinese Cancer Registries included in Cancer Incidence in Five Continents Vol. VIII

5.4 Statistical indices and methods used in cancer statistics

5.4.1 Measurements of cancer burden

Information on cancer patterns and trends provides the basis for evaluating priorities for cancer control in different regions and worldwide. Cancer surveillance needs an accurate evaluation of cancer burden based on the ability to quantify the occurrence and outcome of disease and other cancer-related events in population. Data on cancer occurrence and outcome and the distribution of cancer between different populations and over time are essential to form health policy, quantifying health problems, helping to define priorities for preventive and curative programmes, evaluating their outcome in relation to resource inputs, and establishing causal hypotheses. (Parkin 2001) Quantification of the occurrence and outcome of disease and other health-related events

in population requires a clear definition of the cases or the event of interest (numerator), the population from which the events originate (denominator) and the time frame to which the data refer (time period) (dos Santos Silva 1999). Validity, completeness and reliability are major characteristics required for measurement of health-related events.

Since the emphasis of a cancer control programme is on prevention, the number of new cancer cases each year is the primary measure of interest. In addition to providing insight into the burden on society and on health-care systems, the occurrence and outcome of cancer at various sites are also the basis for estimating the potential for prevention. (Jensen et al. 1991) Various statistical measures are used for assessing the burden of cancer in the population (Jensen et al. 1991, Estève et al. 1994).

Incidence

The single most useful measure for cancer surveillance is the incidence rate, which is an indicator of the risk of disease and can be used to estimate the need for medical attention and treatment facilities. It is the key measure when considering cancer prevention, which aims to reduce incidence, and is of fundamental importance in epidemiological studies. Incidence refers to the number of new cases of cancer occurring in a defined population of disease-free individuals, and the incidence rate is the density of such events in a specified period of time divided by the population's person-time exposure. There are two separate aspects to the calculation of incidence rates: the population estimates which are required as denominators for the rates, and the disease for the individuals. Measurement of incidence requires the identification of all new cancer cases in a defined population through some kind of case-finding mechanism, with record-linkage to ensure that individuals are not confused with events. Cancer registries gather information about the size, nature and distribution of cancer in the community. The major concern of population-based cancer registries is to express the facts about cancer in a standardized way through the calculation of cancer incidence rates for their population, according to histological subtype of cancer, or stage of disease at diagnosis. These rates are used to study the risk of individual cancer in the registry area compared to elsewhere or to compare different subgroups of the population within the registry area itself. Cancer incidence rates are used in addition in planning priorities for cancer control measures such as prevention and early detection programmes, in developing an understanding of the causes and impact of cancer, and in evaluating the effects of control measures. The information on case numbers can also be used for the planning of cancer treatment and care facilities. Using mathematical models of cancer incidence rates for substantive analysis can be useful for hypothesis generation in cancer epidemiology as it allows the investigator to summarize the trends and patterns of cancer incidence through model parameters, which then may then be

tested statistically for the purpose of identifying significant effects of independent variables. Therefore, the accuracy and validity of incidence statistics should be taken into account. Population-based cancer incidence statistics are usually collected by cancer registries, which, unfortunately, are mainly located in developed countries. Furthermore, any data collection system may be inaccurate due to the stage of development of medical and technical knowledge, the availability of diagnostic facilities and also the skill of the investigators. The problem of inaccuracy is best discussed by first considering the separate steps in the chain leading to the production of statistics. (Percy et al. 1981)

Mortality

Mortality indicates only one aspect of cancer burden, although the most serious, or important one. It does not directly reflect the number of persons with disease or volume of health care services, although it is often used as a proxy measure of incidence. For surveillance purposes, mortality statistics have been more widely used than incidence, due to the advantage of more comprehensive coverage and availability. Mortality is the appropriate measure of disease burden in certain circumstances, particularly when the objective of surveillance is to estimate the effectiveness of early detection and treatment programmes. Mortality statistics systems provide basic information on the levels and causes of mortality in populations and can be used to guide decisions regarding health policy and research. Comprehensive mortality statistics require good diagnostic data to be available on deaths that is transferred in a logical, standardized fashion to death certificates which are then accurately and consistently coded, compiled and analyzed. (Percy et al. 1981) However, reliable data on age, sex and cause specific mortality are currently lacking in more than half of all countries (Setel et al. 2005). Major concerns about the quality of mortality data focus on their validity, accuracy, coverage, completeness and also their policy relevance. The validity of mortality statistics is an essential consideration before attempting to interpret any mortality trends in cancer. Several items in the mortality statistics processing chain should be considered when evaluating the validity of mortality: the process of death certification, the classification of disease, and the coding of the cause of death (Alderson 1981). Three criteria, as recently suggested by Rao et al. (2005) could be used for evaluating the validity of the data: content validity, use of ill-defined categories and codes, incorrect or improbable age or sex dependency. The accuracy of mortality statistics depends on the precision of diagnosis, correct completion of the death certificate, validity of the classification of diseases, careful compilation of the tabulations and the consistency of the data with respect to epidemiological expectations (reliability). The characteristics and underlying reasons for under-registration of death vary greatly among countries and also within

each country. Within countries, the completeness of civil registration is known to vary by geographical area and age group. The quality of cause-specific mortality data is affected by limitations of medical knowledge, diagnostic errors, deficiencies of certification, and also coding and other processing errors. The validity of the cancer mortality is also affected by under-registration of deaths. Lack of training in proper certification and insufficient understanding of the uses made of the information provided on the death certificate may cause low quality of death certification. Another frequently encountered problem is physician preferences for certain diagnoses, which may vary from country to country and over time. In many developing countries, a sizable segment of the population lack access to medical care. The situation results in death certificates completed by non-attending physicians, who often have insufficient information to ascertain a diagnosis, and in reports of deaths provided by non-medical witnesses. Furthermore, there are several sources of bias in the study of time trends in mortality. A community-based system – Sample Vital Registration (SVR) – was introduced recently for measuring and monitoring vital statistics that is especially suited to countries lacking comprehensive mortality statistics. A verbal autopsy (VA) method was recommended as one of the techniques of SVR. In VA, a questionnaire is administered to the caregivers or family members of deceased persons to elicit signs and symptoms and their duration, and other pertinent information about the deceased in the period before death. (Setel et al. 2005) As an alternative index of incidence, the accuracy of mortality information collection would also be improved over time, but the improvement has occurred more in the precision of diagnosis than in the number of registered deaths. However, mortality data are frequently of uneven quality and inadequate for the descriptive study of site-specific cancer occurrence. This was considered to be acceptable when a diagnosis of cancer was almost invariably followed by death from the disease. However, improved patient survival for some forms of cancer has led to a divergence of incidence and mortality rates, and an increased risk of multiple primary cancers. In these circumstances, mortality may not be a reliable indicator of cancer incidence. Therefore, cancer mortality is only an indirect (and possibly biased) measure of cancer occurrence. (Percy et al. 1981, Estève et al. 1994, PAHO and WHO 1995, dos Santos Silva 1999)

Survival

Survival statistics deal with the time from diagnosis of cancer to death. They estimate the probability of survival, expressed as time elapsed since diagnosis for individuals within groups. Survival can be used as a means of quantifying the effectiveness of early detection and treatment at the population level, and is thus an important component in monitoring cancer control activities and estimating how many cancer survivors are alive

at any one time. Computation of survival depends upon follow-up information for each individual cancer patient in the group under study, and the calculation of the proportion surviving after different intervals of time. Two related approaches are used to estimate cancer patient survival, usually from randomized clinical trials, observational studies or hospital-based study: the Kaplan-Meier and the life-table method (Estève et al. 1994). In population-based studies of cancer survival comparing survival in groups which are heterogeneous in terms of the risk of death from causes other than a particular cancer of interest, relative survival is often used. Relative survival is generally estimated using life tables, and expressed as the ratio of the observed survival of the patients (where all deaths are considered events) to the expected survival of a comparable group from the general population. The comparison group is matched to the patients with respect to the main factors affecting patient survival, and is assumed to be practically free of the cancer of interest; it is common practically to use nationwide population life-table information. (Dickman et al. 2004) Nevertheless, limitations exist in interpreting survival data: survival from trials or special studies is often biased by patient selection, while population-based survival often lacks detail of stage and treatment information (Sankaranarayanan et al. 1998). The age-structure of the populations studied and reasons for censoring data (loss to follow-up) typically need to be taken into account when comparing survival in different groups, particularly when analysing long-term survival.

Prevalence

Prevalence of cancer is often advanced as a useful measure in cancer surveillance. It describes the number of persons alive at a particular point in time with the disease of interest, and the relative proportion of the population affected by a disease requiring some form of medical attention. Prevalence is intrinsically related to the above three measures, and gives a comprehensive view of the simultaneous effect of incidence, mortality and survival patterns on the cancer burden in a population. In addition, it is claimed that prevalence provides relevant information for planning health services, allocating health resources, administering medical care facilities, designating appropriate research expenditures and assessing the relative burden of cancer with respect to mortality and life quality deprivation. As a measure of cancer burden, it might be more useful to consider as “alive with cancer” those persons still receiving some forms of treatment or being followed up medically, rather than all persons ever diagnosed with cancer at some time in the past, many of whom will be cured and without sequelae. However, the extent and duration of treatment for a given cancer will certainly vary between different populations. It would also be difficult, in practice, to enumerate such patients. For comparison purposes, therefore, prevalence is often

presented as the number of persons still alive after a given number of years following diagnosis. Prevalence is often presented, at least for statistical purposes, only regarding patients alive who have had cancer diagnosed within the last 5 years, which approximates to the period of active treatment and follow-up (Parkin et al. 2005a). Prevalence data derived directly from cancer registries may be biased, since the patients that were diagnosed before the starting of the registry's activity cannot be included in the statistics. The degree of this bias depends on the length of the observation period. Furthermore, differences in prevalence between regions may be due to the variation in risks, health facilities and demographic patterns in local context. For example, populations with long-life expectancy in high-income countries determine a higher prevalence in the areas even for relatively rare chronic disease like cervical cancer (Pisani et al. 2002).

Other indicators

The analysis of the impact of cancer on a population is a complex problem. The above four measures are the most relevant to use for this purpose. However, these variables are seldom analysed jointly in the epidemiological literature, probably owing to their varying availability and reliability, and to the various data sources and collection methods. While mortality data are widely available from official statistics, incidence can be obtained only from specific disease registries or epidemiological surveys, sometimes these also provide prevalence statistics. Survival data are provided by follow-up studies of incidence cases based on clinical records, or on disease registries. Nevertheless, routine registry data is a useful planning tool especially in contexts where with a centralized public provision of health care based on estimated need, rather than private commercial provision of services based on demand and willingness to pay.

In addition to the four measures of disease occurrence and outcome described above, other indices expressing the demand for care are also used in health services planning to develop strategies for service provision. Person-years of life lost (PYLLs) have been widely used in health services planning. PYLL quantifies how many years of the normal lifespan are lost due to deaths from cancer. It refines traditional mortality rates by providing a weighting for deaths at different ages. In burden of disease measurement, health problems of different degrees of severity can be assigned numerical scores on a scale from zero to one. The scores are used to weight life years in calculations of Quality Adjusted Life Years (QALYs) and Disability Adjusted Life Years (DALYs), which quantify the spectrum of morbidity in terms of its duration and severity between onset of a disease and death or recovery. A QALY is an outcome measure that takes into account both the quantity and the quality of the extra life provided by a healthcare intervention – it is the arithmetic product of the life

expectancy and the quality of the remaining years. The more burdensome a disease is to the individual concerned, the lower is its score on the scale. This scoring idea was also adopted in Burden of Disease measurement (World Bank 1993), but using a severity scale where zero represented 'no health problem'. Life years adjusted for health problems are called DALYs. DALYs are the sum of life years lost due to premature mortality and years lived with disability adjusted for severity. The potential value of QALYs and DALYs lies in helping decision makers to compare the severity of different health problems, and thereby to set priorities in prevention and treatment of disease. DALYS has been used as a time-based composite measure of disease burden in both the GBD 1990 and 2000 studies. (Murray et al. 1994, Murray and Lopez 1996, Murray and Lopez 1997, Lopez and Murray 1998, Mathers et al. 2002, Shibuya et al. 2002, Lopez 2005)

5.4.2 Comparing cancer burden between different populations

When comparing measures of cancer risk, outcome, or burden between different regions, population groups or time periods, the effects of differences in age structure of the groups being compared must be taken into account. Comparison of simple crude rates, with no account for the differences in population age-structure may give a false picture. Standardization methods are used for this purpose. Age-standardization, taking a weighted average of category specific rates, can be used to obtain summary measures that are adjusted for differences in age structure. According to the data available, two methods are usually used for calculating the standardized rate. One is direct standardization in which the observed category-specific rates of each population multiple the corresponding category population in the standard population. The other one is indirect standardization for the expected number of cases if the category specific rates were the same as in the standard population. The statistical significance between two rates can be estimated through different methods, such as by roughly testing two age-standardized rates – standardized rate ratio (SRR), using the Mantel-Haenszel test when age-specific rates are available, or through modelling the log-linear function of the relevant covariables. Alternatively, the cumulative rate may be calculated as the sum of age-specific incidence rates multiplied by the widths of the age groups. The cumulative rate provides an estimate of cumulative risk, i.e. the risk that an individual would develop the disease over a defined life span in the absence of any other cause of death. (dos Santos Silva 1999) Furthermore, it is important to know whether the difference between standardized rates, as a summary measure taking a weighted average of category specific rates is due to random variability.

From an epidemiological view, cancer statistics are used to explore the etiological factors and to assess the importance of differences in environment, individual behaviour or variables such as ethnic group, social status, religion, and occupation in cancer causation (Estève et al. 1994). Geographic comparisons are very important in the generation of etiological clues when matching the distribution of certain risk factors. There are often sufficient interesting leads to prompt further study at a local level. A widely used technique is to correlate disease rates and various socio-economic, demographic, or environmental variables, using for analysis small geographic units of aggregates with similar characteristics.

5.4.3 Estimation of cancer statistics based on limited data

5.4.3.1 Methodological approaches for estimating cancer statistics

When estimating and projecting the cancer profile, chronological patterns in incidence or mortality rates depend not only upon the number of cases or deaths but also on the quality of the denominators over time (Alderson 1981). A continuous modification of the population actually at risk throughout the duration of the study must be considered. By definition, except for multiple-primary diseases, a subject is no longer at risk after the occurrence of the event or after the censoring time. A given period of observation of a subject will contribute to the person years in the denominator only if this subject would have been counted in the numerator had he/she experienced the event being studied over that period of time (Estève et al. 1994). However, the population-time at risk is not usually directly measured; it is estimated from the population at the mid-point of the calendar period of interest, multiplied by the length of the period (usually years).

In terms of denominator - measures of numbers of cancer occurrences, since they were usually obtained from the cancer registries, the completeness and validity of the case collecting, coding and reporting directly influenced the accuracy of the cancer statistics estimation. Measures can be calculated for the whole population, as crude measures, or separately for certain subgroups of the population, as stratum-specific measures. The Poisson assumption for the number of cancer cases has been shown to be superior to methods that assume a normal distribution for the cancer rates, using a Finnish example (Dyba and Hakulinen 2000). A common assumption, therefore, is to treat the numbers of cases or deaths of cancer in strata formed by age and period as independent Poisson-distributed observations.

With differences of data availability and quality, the methods used and the accuracy of estimates of cancer burden will vary between populations. For countries with only regional cancer registration or lower quality of registration data, different

methods have to be considered, according to the local information set, exploiting the empirical relationship between different statistical indices in cross-sectional data and applying different models for selected cancer registry areas extrapolated to the national profile.

If sufficiently detailed survival data are available from a population-based cancer registry, it is technically possible to estimate the empirical relation which links incidence, mortality and survival, and then to use this observed relation to estimate cancer incidence or mortality. For a given cancer site and age group, mortality (M) is the product of incidence (I) and the probability of dying from the disease

$$M = I \times [k - S_j]$$

where S_j is the relative survival at year j of follow-up and k is a constant depending on j . When 5-year relative survival probabilities are used, the constant k tends to be very close to 1. (Estève et al. 1994, Pisani et al. 1999) Similar mortality prediction technology, which is based on the prediction of incidence and on those made for cancer patients' survival has also been used when estimating of the cancer mortality burden in five Nordic countries, where dependable and complete cancer incidence registration systems have been established for decades, from the 1940s in Denmark and the 1950s in the other Nordic countries (England et al. 1995), also been successfully and more detailed used in certain particular settings such as in estimating breast cancer statistics in Italy (Capocaccia et al. 1990).

Nevertheless, population-based survival is only available for a few populations or certain ad hoc studies on cancer. Furthermore, with genuine differences in patient survival due to stage at diagnosis, access to care and quality of treatment, possibly artifacts of data collection and death certificate only (DCO) cases included in registration data while excluded from survival analysis, published survival data may not be suitable for direct use in populations other than those from which they came.

Therefore, simpler but not less reliable methods have often been used, such as an empirical approach based on a regression equation estimated from similar countries or geographical regions where both incidence rates and mortality data are known. One such method is based on incidence-mortality ratios when evaluating national cancer profile from regional cancer registration data, from mortality to estimate incidence, or vice versa. It was first used by Doll (1960) when estimating incidence rates for cancer of certain sites in populations of a variety of countries. Similar methods for estimating national cancer statistics have also been used to estimate national incidence or mortality in GLOBOCAN and in Europe for those countries without full coverage by cancer registration schemes (Parkin et al. 1988, Jensen et al. 1990, Estève et al. 1993, Ferlay et al. 1999, Parkin et al. 1999, Ferlay et al. 2001, Ferlay et al. 2004).

The underlying assumption involved in the above method is that the ratio of incidence to mortality (I/M) is the same in the cancer registry populations used to estimate the model parameters (I_R/M_R) as in the larger national populations (I_N/M_N). It is important that the source of the mortality data is the same for M_R and M_N , with only the geographic scope differing. The national incidence (I_N) can then be estimated by applying a set of age, sex and site-specific incidence mortality ratios (I_R/M_R) obtained from the aggregation of representative cancer registry data, to the corresponding country's national mortality (M_N):

$$I_N = M_N \times I_R / M_R$$

Estimates of the I_R/M_R were usually obtained from log-linear models with Poisson error for the numbers of incident cases in the pooled regional registries offset by the numbers of deaths in the same registries with terms for sex and age. Such terms for sex and the polynomials of age up to the low degree were introduced on the basis of statistically significant changes in the goodness-of-fit of the models. Successive polynomial terms for age were included until two consecutive terms failed to provide a significantly improved fit to the data, or the maximum of five was reached. In general, this approach worked well, but practical problems were encountered when trying to fit models for cancer sites for which incident cases were associated with few deaths. To achieve a greater degree of stability in the resulting estimates, Bray et al. (2002) used average mortality estimates for the latest 3 years available and weighted aggregation relative to population size.

The accuracy of the original incidence and mortality data has been considered by Black et al. (1997) when using this method to estimate the cancer incidence and mortality in the European Union in 1990. Inaccuracy of the mortality rates (for example, due to incomplete ascertainment, or miscoding) is, in theory, not important in the use of the model to accurately predict incidence, providing that the accuracy is the same in the area included in the model ('equation area') and the area for which the estimation is being made ('estimation area'). Thus, although the I/M ratios themselves may not be an accurate reflection of reality (of cancer survival, in fact), this is not necessarily of any consequence to the model prediction of incidence. It is therefore important that both sets of mortality data are derived from the same source. However, any systematic differences in completeness or accuracy of death classification between the areas would distort the results. If the I/M ratio is higher in the 'estimation area' than in the 'equation area', a spuriously low number of incidence cases will be obtained. Such a situation may arise if survival following a given cancer is better in the estimation area. A similar situation may arise if more deaths in the 'estimation area' are classified to unknown and unspecified causes or cancer categories. On the other hand, an overestimation of the cancer incidence for the population in the 'estimation area' will occur if the I/M ratio in

the ‘equation area’ higher. This may happen if survival is poorer in the ‘estimation area’, or if the mortality rate is spuriously high – for example as a result of misclassification of metastases as primary cancers. Differences in the age-specific I/M ratios would lead to unpredictable distortions of the incidence estimates. In addition to such systematic differences, random fluctuations are likely to affect numbers based on few observations, although the smoothing effect of the regression models used for the age-specific I/M ratios would tend to reduce the variations in the statistics’ accuracy.

Furthermore, by model fitting, if the actual variance of the observed number of cases or deaths is larger than expected under the Poisson assumption, the model is said to have over-dispersion or extra-Poisson variation. This is not uncommon when counts are large, for instance when data at national level are studied. One way to cope with this problem is to modify the fixed relationship between the mean and the variance in the Poisson distribution by including a proportionality constant, called the ‘heterogeneity factor’. The factor can be estimated from the deviance of the most complex model prepared to consider, provided that it contains all appropriate explanatory variables. From a theoretical point of view, the considerable extra-Poisson variation existing in the above estimation model would generally make the results conservative. However, a careful inspection of the age-specific fitted values indicated that the overdispersion has no practical influence, and is also theoretically true, on the validity of the average relation obtained. The extra-Poisson variation does not bias the estimates of the coefficients of the regression equation. Nevertheless, with all the above potential errors or bias, the estimation of cancer incidence to determine the burden of cancer in a population is only a surrogate for measuring incidence by means of systematic cancer registration.

5.4.3.2 Cancer estimation in practice

Different methods have been used in the latest released GLOBOCAN2002 database, depending on the degree of detail and accuracy of the cancer information available in each country. For estimating the sex, age-specific incidence rates of cancer for a country, six different approaches were used, depending on the availability of data: national incidence data available, estimating from national mortality data (assuming that the logarithm of the incidence rate of a given cancer can be expressed as a linear function of the logarithm of mortality in each sex-age group and projecting the incidence-mortality ratios from qualified registries (data of those cancer registries accepted in CI5-VIII) or sample data to national level), estimates derived from local registries’ data in a country, using any available data on the relative frequency data of different cancers or directly using other countries’ data if there are no data available in

the country. (Parkin et al. 1999, Parkin 2001) Cancer mortality statistics by cause were derived from one of three possible sources: WHO Mortality Database (containing national or representative mortality data), the mortality data from cancer registries, or *ad hoc* publications of mortality data from a local source. For countries with no mortality data or data known to be of poor quality, mortality was estimated from country or region-specific incidence and survival data (Pisani et al. 1999). Sources of population-based survival, shown as relative survival, used for the developing countries were from the ‘Cancer Survival in Developing Countries’ project (Sankaranarayanan et al. 1998) or local research projects.

Mortality data are available for all countries of the EU, but nationwide cancer incidence rates are recorded only for a few of them. Jensen et al. (1990) used methods similar to those of GLOBOCAN to estimate the number of incident cases in 1978–1982 at a given age from the number of deaths at the same age in 1980–1984 for the EU and its member countries. A generalized log-linear model with Poisson error was used and the age included in a polynomial of low degree determined by significance testing. The latest estimation of cancer incidence and mortality for the whole of Europe in 1995 was made by Bray et al. (2002), using a similar methodology, incorporating the many local and national cancer data sources available.

Different measures are used in the SEER*Stat program to describe the cancer statistics in USA, shown in different sessions in the programme (Case-Listing, Frequency, Rate, Survival, Limited-Duration Prevalence and MP-SIR) with numerous statistical methods as a means to logically organize these statistics (Testa and Meigs 1980, Wingo et al. 1998, Kim et al. 2000). An incidence-based mortality (IBM) rate is used based on the high quality of population-based cancer registries data (including follow-up information) that allows the mortality event linked to factors identified with the disease at the time of diagnosis (e.g. tumour stage and grade). Nevertheless, it is noted that factors like lead-time bias can influence the IBM analyzing result, cautions need to be taken when interpreting the results. Companion software with different statistical models are provided by the SEER programme, such as Joinpoint Regression Program (using joinpoint models for calculating the population-based cancer trends in USA), DevCan (computing the probability of developing or dying of cancer from birth or conditional on a certain age through calculating incidence and mortality rates in the standard areas of SEER).

5.4.4 Trends in incidence and mortality

The occurrence and outcome of cancer may vary over time. Information on the historical evolution of trends in incidence or mortality can generate etiological

hypotheses or confirm suspected associations between risk factors and disease. The observation of time series can also be seen as an instrument for epidemiological surveillance of the population with the aim of detecting new risk factors. The study of time trends is of particular interest in the evaluation of primary prevention (involving reduction in exposure to risk factors) and of secondary prevention (screening) (aiming at reducing mortality). From the public health viewpoint the observation of changes in risk in the recent past leads naturally to a desire to predict the future development, in order to determine budget priorities and plan necessary services.

Ideally, although deaths and years of life lost through cancer have been used widely in the context of prevention, from an etiological point of view, up-to-date information on cancer incidence is clearly the most relevant parameter when studying the historical evolution of cancer. Time series of incidence partially reflect progressive improvements in the registration rate, whether resulting from the development of diagnostic techniques or improved reporting systems for the registry. The newer the registry, the stronger this effect is likely to be. Unfortunately, trends in incidence can be only studied in a few countries, such as the Nordic countries, where there are complete registers of newly diagnosed cases. In addition, in many countries, results can rarely be generalized because registries often cover subpopulations chosen by circumstance. As indicated before, mortality is often used as an alternative measurement due to its wider availability.

Changes over time in cancer incidence or mortality represent effects occurring on three different scales (effects): age effects, calendar time (period effects) and date of birth (cohort effects). These can be examined separately in an age-by-calendar period two-way Lexis-diagram in which the diagonals represent successive birth cohorts (Estève et al. 1994). Age is obviously the most powerful determinant of cancer risk, and cancer incidence rates usually increase with age. The rates may also differ over calendar time and between birth cohorts; however, often their relative contribution cannot be measured directly. Period effects are influenced by an immediate or fixed-delayed change in the incidence rates for all age groups, due to changes in classification criteria, the availability of new diagnostic tests or the introduction of a powerful carcinogen or a screening intervention. Cohort effects imply changes in incidence rates from one generation to another that is consistent across age groups.

Observed time trends should be evaluated in the context of the problem under study. Sometimes it is sufficient to describe long-term trends, while more often interest is focused on variation over a more limited time period, in particular the recent past, when the goal is to predict new directions of the phenomenon. New techniques, mainly based on mathematical modelling have been developed to distinguish between the different effects or factors that underlie changes in rates. Apart from the simple

description of changes in cancer rates over time, the study of trends should involve the search for models that can describe observed data via plausible hypotheses about the causes of observed changes. In such approaches, the relevant components of the time trend can effectively be separated from the random or systematic, allowing a more complete interpretation of the observed data. But, the difficulty of separating meaningful variations (random fluctuations) and the cohort and period effects has restricted the use of models. Nevertheless, various descriptive methods and mathematical models have been widely used during last decades, including graphical displays, the overall or mean annual percentage rate of change in the logarithms of age-standardized or age-specific rates, as well as modelling of age, period and cohort effects.

A good graphical display of time trends is invaluable as the sole means for comparing trends between many countries or regions. It relies on the visual inspection of plots of rates to describe the changes in the patterns, and to analyse temporal trends, especially through calculating a constant average annual percent change over time. Devesa et al. (1995) indicated that a graph should be designed and studied with care, clearly reflect the truth, convey information, and make a point without overemphasising. Graphical presentations can be improved by choosing an arithmetic or logarithmic scale according to the research question being addressed, by selecting scaling ratios of the axes that allow one to detect specific rates of change, and by using uniform scaling ratios to facilitate comparisons between graphs.

The annual percentage change in age-standardized or age-specific rates is often used combined with graphical display. Such annual changes in age-standardized rates, however, are strictly interpretable only when the effect of calendar time on cancer risk is multiplicative, i.e. when the age-specific incidence or mortality curves for successive calendar periods are parallel on a logarithmic scale. Trends observed in age-standardized rates may be misleading when the time trends in different age groups are significantly different from one another. Examination of age-specific rates alone may suggest the existence of a cohort effect, for example with a rapid increase in risk of younger ages and progressively smaller rates of increase, or even decline in successively older age groups.

Nevertheless, a graphical display, even with an annual change described, is inadequate for presenting the quantitative changes of rates. Quantitative and comparable estimates of trends can be obtained via statistical models based on objective criteria for selecting the best description of the data, and including significance tests of whether the observed trends are real or random. Modelling of age, period and cohort effects has been used increasingly as a widely accepted method providing a succinct and interpretable summary of the data. Three separate models are often used based on the

data available and usually fitted by weighted least-squares or by Poisson maximum likelihood methods. (Clayton and Schiffers 1987a, Clayton and Schiffers 1987b, Coleman et al. 1993)

1. The linear age-period model for the logarithmic of the expected value of the rate $E[M_{it}]$ is presented as

$$\ln E[M_{it}] = \alpha_i + \beta t,$$

where β is the (constant) change in log-rates from one period to the next, α_i is the fitted age-specific rates in age i in the reference period coded zero. In this model, the age-specific curves in successive calendar periods are according to the model parallel on a logarithmic scale.

2. The corresponding age-cohort model is similarly formulated by

$$\ln E[M_{ic}] = \alpha_i + \beta_c c,$$

where β_c is the (constant) change in log-rates from one cohort to the next and α_i is the fitted age-specific rates in age i in the reference cohort coded zero. The age-specific curves for successive birth cohorts are according to the model parallel on a logarithmic scale.

3. The corresponding age-period-cohort model depends on age, calendar period and year of birth:

$$\ln E[M_{itc}] = \gamma_i + \beta t + \beta_c c.$$

However, it cannot be uniquely estimated from empirical data (Clayton and Schiffers 1987b). In order to overcome this so-called non-identifiability problem, a special single parameter, the ‘drift’ has been introduced for a variation over time which not specifically attributable either to period or cohort effects. It is the sum of the linear components of increase in the logarithm of the rate due to period and cohort effects combined, it is identifiable and may be estimated. For data sets best fitted by a model including only linear terms for period or cohort, thus for which the incidence or mortality rate is changing at a constant rate over time, the ‘drift’ is the same as the slope of the line plotted in a graph of the logarithm of the incidence or mortality rate against time (calendar period or birth-year interval). According to the relationship birth year + age = period ($c + i = t$), the same drift models can be introduced, for example, either as

an age+period-drift model: $\ln E[M_{it}] = \gamma_i - \beta_c i + (\beta + \beta_c)t,$

or as an age+cohort-drift model: $\ln E[M_{ic}] = \gamma_i + \beta_c i + (\beta + \beta_c)c$

with net or total drift parameter $\beta + \beta_c$ held constant in any (age+period-drift+cohort-drift) model. (Clayton and Schiffers 1987a, Clayton and Schiffers 1987b, Coleman et al. 1993)

Non-linear terms P_t and C_c may be added also for period and cohort giving for instance for the age+period-drift model:

$$\ln E[M_{it}] = \alpha_i + \beta i + P_t + C_c$$

Data used in above models are often arranged in a two-way table by age group and calendar period, which are mostly subdivided into 5- or 10-year intervals. It is indicated that one of the disadvantages of those models is the loss of information by data aggregation and with problems of estimating interactions in the two-way layout without replications. Heuer (1997) showed that using restricted regression splines for yearly data is a more accurate method for curve estimation for the nonlinear trend changes compared with the traditional approach for aggregated data and a simple way of modelling interactions between the time variables.

The method used by Coleman et al. (1993) for analysing the trends in cancer incidence and mortality was based on polynomial age-period-cohort models, fitted as generalized linear models with Poisson-distributed errors and logarithmic link. Polynomial functions were used to obtain effective smoothing of the age, period and cohort effects. In order to avoid producing spurious results when the data are markedly irregular, the degree of the polynomial used in each model was kept low unless there was a strong argument to the contrary, and extrapolation beyond the range of the observed data was kept to a minimum. It was indicated that efficient smoothing is especially important for the study of cohort effects since the calculation of the cumulative risk for the older and younger generations represented in the data implies a considerable degree of extrapolation. Coleman et al. (1993) did not extrapolate any non-linear period effects (on the logarithmic scale used).

If the interest is in identifying and describing the occurrence of changes in recent trends in distinct periods of time, a newly developed method – the joinpoint regression model - can be used. For the observations $(x_1, y_1), \dots, (x_n, y_n)$, where $x_1 \leq \dots \leq x_n$ without loss of generality, this model is written as

$$E(y/x) = \beta_0 + \beta_1 x + \delta_1 (x - \tau_1)^+ + \dots + \delta_k (x - \tau_k)^+$$

where the τ_k 's are the unknown joinpoints and $a^+ = a$ for $a > 0$ and otherwise 0. For a model with k joinpoints, the i th response y_i is formulated as:

$$y_i = \beta_0 + \beta_1 x_i + \delta_1 (x_i - \tau_1)^+ + \dots + \delta_k (x_i - \tau_k)^+ + \varepsilon_i^{(k)} = \mu_i^{(k)} + \varepsilon_i^{(k)}$$

where $\varepsilon_i^{(k)}$ is the error and $\mu_i^{(k)}$ is defined implicitly. The program starts with the minimum number of joinpoint (e.g. 0 joinpoints, which is a straight line) and tests whether more joinpoints are statistically significant, which must then be added to the model (up to a user defined maximum number). The tests of significance use a Monte Carlo permutation method. In the final model, with the best fitting joinpoints where the rate changes significantly, each joinpoint informs of a statistically significant change and an estimated annual percent change (EAPC) is computed along with its 95% confidence intervals. The model expresses the logarithm of the expected number of cases or deaths as a linear function of the factors and follows a Poisson distribution. Significant changes include changes in direction or in the rate of increase or decrease. (Kim et al. 2000, NCI – SRAB 2005)

5.4.5 Projection of cancer statistics

Projection of the occurrence and outcome of cancer are essential for both administrative and scientific purposes. For administrative planning, an accurate future prediction would allow finite resources be optimally allocated to the core elements of cancer control, such as primary prevention, screening and early diagnosis, treatment, rehabilitation and palliative care. In a scientific context, predictions are made for evaluating the success of intervention or early detection programmes. (Hakulinen and Dyba 1994) Estimations and projections can be used as a baseline in monitoring the changes of cancer incidence or mortality rates over time.

A projection that estimates the future demand for health services may be based on previous trends, and therefore is a composite of numerous factors related to past progress. It is clearly important that the expected trends or numbers be as accurate as possible or close to the current or future observations. Nevertheless, predictions, as with temporal analyses, involve difficulties at the data collection, data presentation and data interpretation stages of a study and with the lengthy and variable latency period between carcinogenic exposure and development of some cancers (Hakulinen 1996), predictions obtained from simple extrapolations may be inaccurate. A clear and true picture of the cancer situation in the country can best be achieved with a complete nationwide or otherwise representative cancer registration system. Nevertheless, it is hard work with lack of qualified cancer registration data in many countries, especially in developing areas.

Ideally, if both the prevalence of an exposure and its effect on a specific cancer can be quantified, information on etiological factors can be used as a basis for making predictions. Although hypothetical rates of lung cancer in future periods have been predicted by taking into account the uptake and cessation of smoking at the population level (Hakulinen and Pukkala 1981, Brown and Kessler 1988, Yamaguchi et al. 1992), unfortunately, a precise understanding of the factors that drive most cancer trends is currently lacking. Further, data are seldom routinely available at the requisite level of detail – by sex, age group and time period – and in any case, few associations between a single risk factor and the onset of cancer are strong enough to be modelled directly. In addition, the definitions, diagnostic criteria and facilities change over time, making or masking trends in the occurrence or outcome of cancer. Therefore, in most instances, for better or worse, predictions are based on the extrapolations of past time trends.

Numerous discussions have focused on the sensitivity of the predictions based on model extrapolations into the future. Hakulinen (1996) pointed out that if the model is not correct, the error would rapidly increase in the extrapolation. Nevertheless, even when an appropriate prediction model has been chosen, the extrapolated results will still

have random errors from both historical and future observations. Model assumptions that held true with historical data may not necessarily hold true in the future. Furthermore, a good fit does not automatically guarantee a good prediction. This is particularly true for more complicated models containing several parameters. It would not be good to include higher order terms, which, although possibly improving fit, increase the uncertainty of the predictions in a dramatic way. The principle of parsimony is crucial in the context of prediction.

Various modelling strategies are used for predicting future cancer patterns based upon past trends. The reliability of prediction depends directly on the choice of model and the variables included in it. In estimating the numbers of future cases or deaths for public health purposes, it is very important to make a distinction between changes due to the demography of the population under study (population size and age structure), and changes due to an evolving risk pattern over time. The evolution of the risk of cancer should ideally, as mentioned above, take into account changes in the prevalence of exposure to etiological factors. In practice, when information of risk factors is not sufficient for prediction modeling, they are represented by surrogate variables such as age, period of observation and year of birth (cohort).

The choice of statistical model for extrapolating disease burden into the future is a crucial part in the process of prediction. The predictions depend not only on the type of model, but also on the method of extrapolating the model components. Many complex models, such as polynomial age-period-cohort models of complicated mathematical form or with many parameters, may fit the data within the range of observations. For prediction purposes, the age-period-cohort model usually involves estimating the underlying age- period- and cohort specific trends and projecting them into the future using a Poisson regression model. These multiplicative models may yield rates that grow exponentially with time since the predictions are calculated based on the assumption that current trends will continue similarly in the future. For the prediction period, it is not realistic to assume that the effect of current trends will continue to the same extent throughout a long period, such as 25 years, already far beyond anything that has realistic practical meaning.

The simplest way of making a prediction is to extrapolate a linear model, which assumes no change of underlying trends in the basis of prediction. Hakulinen and Dyba (1994) suggested that a constant absolute increase in incidence (in a population age stratum) would be chosen for cancers with an increase in the incidence, and a constant proportional decrease in incidence for cancers with a decrease in incidence. These simple linear models on an arithmetic or logarithmic scale seem reasonable and avoid exponential growth in rates and linear decrease, leading to implausibly low or negative predicted incidence rates (Hakulinen and Dyba 1994, Dyba et al. 1997, Dyba and

Hakulinen 2000). Following models are often used to predict the expected rates of the cancer, $E[M_{it}]$, in age group i and year t , is, using of the Poisson-distributed counts, in strata defined by sex, age and time period:

1. assuming the same constant proportional (percentage) change for all age groups in a log-linear model:

$$\ln E[M_{it}] = \alpha_i + \beta t$$

(same as the age-period model considered above).

2. assuming different proportional changes for different age groups in a log-linear model:

$$\ln E[M_{it}] = \alpha_i + \beta_i t$$

3. assuming different linear changes on the absolute (arithmetic) scale:

$$E[M_{it}] = \alpha_i + \beta_i t$$

4. assuming proportional effects for different age groups, but with the constraint that each age-specific absolute change of incidence is proportional to the corresponding age-specific baseline rate (α_i) which gives a smoother set of prediction and narrower age-specific prediction and confidence intervals:

$$E[M_{it}] = \alpha_i (1 + \beta t)$$

where α , α_i , β and β_i are unknown parameters.

In practice, because of the long history of efficient population-based cancer registry systems, the prerequisites for predicting cancer incidence statistics are most favourable in the Nordic countries (Denmark, Finland, Iceland, Norway and Sweden). Wiklund et al. (1992) provided a rough prediction of cancer mortality in the Nordic countries in 2005. It was based on existing and easily accessible data concerning population, cancer incidence and mortality, and performed with the American software CAN*TROL, using a mathematical model to calculate the effect of various interventions on future cancer incidence, prevalence and mortality. In France, Remontet et al. (2002) recently provided the estimates of national cancer incidence and mortality trends from 1978 up to 2000. An age cohort model was used for modelling the observed incidence and mortality data in the population covered by cancer registries. Then an estimation of the incidence/mortality ratio was obtained from these models, and applied to the mortality rates estimated from an age-cohort model for the entire French population. The person-years of observation were calculated cohort-wise from census data provided by the national institute of statistics. The age-cohort model has also been used to estimate future lung cancer incidence trends in 2015 in United Kingdom (Parsons and Somerville 2000).

The Association of the Nordic Cancer Registries (ANCR) was also requested to conduct a complete description and analysis of cancer patterns in the Nordic region of Europe. A series of publications using more advanced and complex methods to achieve

this goal was started with Engeland et al. (1993) predicting cancer incidence up to the year 2012 based on observed cancer incidence rates from 1958-1987. The predictions are based on age-period-cohort models, but a *power* transformation (link) was used instead of the *log* link for Denmark, Finland, Norway and Sweden:

$$E[M_{it}]^\rho = A_i + \beta t + P_t + C_c,$$

where $E[M_{it}]$ is the incidence rate in age group i in calendar period t , ρ is a constant power between 0 and 1, β is the drift parameter, A_i is the age effect for age group i , P_t is the non-linear period effect of period t and C_c is the non-linear cohort effect of cohort c . The predictions were based on the assumption that cohort-specific patterns would continue to hold during the prediction periods, with some reduction in the drift in the last periods. A model with $\rho = 0.2$ was chosen, empirically, which was close enough to the multiplicative model to give a good fit to the observed data, and at the same time the exponential growth in the predictions was levelled off.

This *power* prediction model was also used by Møller et al. (2002) for predicting cancer incidence in Nordic countries up to the year 2022, based on observed cancer incident cases in 1968–1997. Instead of using average drift to analyse trends, they used the trend from the ten most recent years if there was significant curvature in the rates over the last 20–30 years. The numbers of new cancer cases in five periods from 1998 to 2022 were predicted. When predicting incidences of cancers in breast and prostate, the effects from mammographic and PSA screening have also been taken into consideration.

6. Aims of the Study

The aim of this study was to estimate the current cancer burden (incidence and mortality) at the national level in China, using the most accurate available cancer information on both incidence and mortality, to give reference values for the cancer prevention programme in China. In order to achieve this, several studies are summarized in this thesis:

1. An investigation of the cancer registration situation in China through a national survey;
2. An investigation of the quality of the different cancer mortality data sets, and in particular, how representative they are of the overall Chinese population;
3. A review of cancer mortality time trends, including differences in trends by sex, age group, site, and in urban and rural populations, and speculating on the possible reasons;
4. The estimation of national rates of cancer mortality, for the major cancers
5. Using the time trends to project cancer mortality and incidence in 2000 and 2005, as a guide to priorities for cancer control in China.

7. Materials and Methods

7.1 Data sources

- 1) Data on the Chinese population by age, sex and area (urban and rural residences) for 1990 and 2000 were obtained from two national censuses (PCO-NBSC 1993, PCO-NBSC 2002); Projections of the population by age and sex in 2005 were obtained from the UN Population Division – world population the 2002 version (UN-PD 2003); the estimated ratios for urban to rural of the total population in 1995, 1998, 1999 was also from UN released data (UN-SD 2002; UN-PD 2002);
- 2) Cancer mortality information, by sex, age, area and sites, was derived from the second national retrospective survey of deaths in 1990–1992 (Li et al. 1996, Li et al. 1997b, Zhou et al. 1997); the data submitted from the Center for Health Information and Statistics (CHIS) during 1991 to 1999 (WHO 2005b); the Disease Surveillance Point (DSP) mortality system during 1992 to 1998 (NDSPTS 2002);
- 3) Cancer incidence and mortality information for the period 1993–1997 was extracted from seven population-based cancer registries in China whose data had been published in Cancer Incidence in Five Continents, volume VIII (Parkin et al. 2002).

7.2 Methods used in the study

(1) **Population estimation and projection:** (Paper III)

The only available comprehensive data on the Chinese population are derived from censuses, which taken in 1964, 1982, 1990 and most recently, in 2000, detailing the sizes of the population by age, sex and area (urban and rural residences). In this study, the national population data were derived from the censuses in 1990 and 2000 (PCO-NBSC 1993, PCO-NBSC 2002). Age-sex specific population data for year 2005 was obtained from the UN estimation (UN-PD 2003), however, the data were not specified by area (rural / urban). The annual rate of change of the proportion of the total population classified as rural during 1995 to 1999 was then calculated from the UN estimates (UN-SD 2002) and 2000 census data (PCO-NBSC 2002). The exponential of this change was applied to

estimate the proportion of the rural population in 2005 (and, by subtraction, the urban population). This proportion (all ages by sex in rural area) was used to obtain the proportion in individual age groups, by multiplying the ratio of %rural in the age group concerned to the %rural of all ages, based on the corresponding population distribution of the 2000 census. These proportions were then applied to predict the numbers of the age-sex-area specified Chinese population in 2005, based on the UN estimation.

(2) **To investigate the cancer registration practice in China:** (Paper V)

A two-stage survey was conducted in 2002 in China – at provincial and cancer registry level. At the provincial level, a contact person was identified in the Health Department of each province, who was responsible for identifying all existing cancer registries in the province, and for the administration of the registry level questionnaire. At the registry level, each cancer registry was asked to fill in a questionnaire about its cancer registration practices. Five major aspects were included in the questionnaire: background; sources of data (incidence and mortality); information recorded (data items, coding, follow-up, data checking and/or linkage); data management (data entry, processing of data, checking, linkage, quality control) and output (reports and research). An overall profile of current cancer registration practices in China was then established, based on the information in the questionnaires.

(3) **To compare the validity of mortality data from available data sources in China:** (Paper II)

The distributions of the population at risk, by age, sex, and urban/rural ratio from the second national mortality survey 1990–1992, and from CHIS and DSP mortality datasets in 1992 were compared, using the 1990 census population as a standard. Standardized rate ratios (SRR),

$$SRR = ASR_1 / ASR_2,$$

were used to show the differences between the area-specific age standardized rates (ASR), referring to the world standard population, in the DSP and CHIS mortality data, and those recorded in the second mortality survey 1990–1992. Estimates of the national age-standardized mortality rates (rural and urban areas combined) by sex, for 9 cancers and all cancers combined, were prepared, based on the population distribution of the 1990 census population. Comparison of changes in age standardized cancer mortality between 1992 and 1998, using the DSP and CHIS datasets, was performed for four cancer sites (stomach, lung, breast, cervix). The estimated annual percentage changes (EAPC) for each sub-population were calculated by linear regression of the *log* (mortality rate) for both datasets.

(4) **To investigate time trends in cancer mortality during 1987–1999:** (Paper I)

Using the cancer mortality dataset that had proved to be most representative of the national mortality profile from the comparison in step (3), age-adjusted annual mortality rates over the 13 years (1987–1999) were calculated by the direct method, standardized to the national census population in 1990, and trends by site, sex, age group (15–34, 35–44, 45–54, 55–64, 65–74, 75+ (plus, for leukaemia, age groups 0–4 and 5–14)) and area (urban, rural) plotted on a logarithmic scale against year. The joinpoint regression model was used to estimate the changes in the age-standardized mortality rate, shown as EAPCs with its 95% confidence intervals, through expressing the logarithm of the expected number of deaths as a linear function of the factors and following a Poisson distribution (Kim et al. 2000; Fernandez et al. 2000). Age effects were analysed based on the Poisson regression model for the age specific mortality rates by cancer, sex and area. The annual percent change in the trend for each age group was also evaluated. (Breslow 1984, Coleman et al. 1993, Estève et al. 1994)

(5) **To estimate and project the national cancer mortality in 2000 and 2005:** (Paper III)

Based on the Poisson regression model, the EAPCs in the mortality rates for common cancers, as well as all other cancers combined, by area (rural/urban), sex and age group from 1991 to 1999 were estimated for the same cancer mortality dataset as the one used in step (4). The changes were then applied to the mortality rates observed in the second national mortality survey in 1990-1992, to estimate the age-specific mortality rate at the national level for the corresponding cancers from 1991 to 2005 based on a log-linear model

$$\log E[M_{it}] = \alpha_i + \beta_i t .$$

Here α_i is the baseline age-specific mortality rate (when $t = 0$ e.g. year 1991) and β_i is the observed EAPC in the dataset for age group i .

The validity of choosing above prediction model was tested in a sensitivity analysis, by comparing the results with those obtained using three other prediction models:

$$\log E[M_{it}] = \alpha_i + \beta t$$

$$E[M_{it}] = \alpha_i + \beta_i t$$

$$E[M_{it}] = \alpha_i(1 + \beta t) .$$

To validate the choice of prediction model, each of the four models was fitted to compare observed and expected numbers of deaths, age-specific and age-standardised rates of cancers of the lung (in rural females), breast (in rural females), oesophagus (in urban females) and cervical cancer (in urban females) in the actual data from the dataset chosen in step 2 in 1999. The differences between

observed and expected numbers of deaths, age-specific and age-standardized rates were compared.

It was assumed, due to sparse numbers of cancer deaths, except for leukaemia, that mortality remained constant in the youngest age group (less than 15 years old). Combined with the population in the corresponding year, the age sex area-specific expected numbers of deaths, for each cancer were calculated. ASRs were calculated for 1991, 2000, and 2005, adjusted by the weights of the world standard population.

The two components of the change in the number of deaths between 1991, and 2000 or 2005, were presented separately: changes due to risk calculated as the difference between the estimated future deaths and the number that would have occurred if the mortality rates had remained the same as in 1991; and changes due to population growth and ageing which obtained as the difference between this latter figure, and the original number of deaths (in 1991). (Engeland et al. 1993)

(6) **To estimate and project the national cancer incidence in 2000 and 2005:**
(Paper IV)

The numbers of national cancer incident cases (I_{Nij}), where N refers to the national data, i to age group and j to sex, were estimated for ten common cancers (nasopharynx, oesophagus, stomach, colon-rectum, liver, lung, female breast, cervix, leukaemia and bladder), using the estimated and projected national cancer mortality rates in 2000 and 2005, derived from step (5), together with cancer incidence and mortality information for 1993–1997 from seven population-based cancer registries in China whose data were published in CI5-VIII (Parkin et al. 2002): Beijing, Shanghai, Wuhan, Qidong, Jiashan, Cixian and Changle (the latter two were only used for estimating the incidence for cancers in the oesophagus and stomach). Based on the methods used in earlier publications by Jensen et al. (1990) and Bray et al. (2000), the estimates were as the product of the estimated national mortality (the age-sex-specific numbers of deaths) (M_{Nij}) and the incidence-mortality ratios from the aggregated cases and deaths from above cancer registries (I_{Rij}/M_{Rij}), here R refers to registry data, weighted by the reciprocal of the square root of the average population at risk and fitted by a generalized log-linear model with Poisson error.

$$I_{Nij} = M_{Nij} \times I_{Rij} / M_{Rij}.$$

The GLIM programme (Francis et al. 1993) was used to fit the models including explanatory terms for sex and age. For the category ‘all cancer sites’, estimates were obtained by summing the number of the ten common cancer cases and the ‘all other cancers’ category.

The validity of using the aggregated cancer registries' ratio of incidence to mortality I_{Rij} / M_{Rij} was examined in a sensitivity analysis. This used the estimates for oesophageal and stomach cancer incidence (which had data from 4 rural registries (Cixian, Changle, Jiashan and Qidong (Rural (4))), while for all other sites, only 2 rural registries (Jiashan and Qidong alone (Rural (2))) were available), and compared the resulting estimates with those from two stratified analyses. The latter used three datasets of I_{Rij} / M_{Rij} : one based on three urban registries in Beijing, Shanghai and Wuhan (Urban), one from Rural (4), and one only from Rural (2). The combinations of urban with above two rural strata aggregated I_{Rij} / M_{Rij} were compared for the validity of the estimation. It was assumed that Urban + Rural (4) would be the best, while at the same time, seeing if Pooled (7) was very different or not. For the national incidence profile estimating, since these data were only available to estimate two cancer sites – oesophagus and stomach, for most other sites, Urban + Rural (2) or Pooled (5) were used, so the interest was in which would produce the most valid estimate.

The “all other cancers” category was partitioned according to the proportions from the registries' data. Combined with population data, the estimated or projected age-specific incidence rates in 2000 and 2005 were obtained. Cumulative risks (up to age 74) were used as the net risk

$$Cum.Risk(\%) = 100 \times [1 - \exp(-(r_{0-44} \times 45 + r_{45-54} \times 10 + r_{55-64} \times 10 + r_{65-74} \times 10))]$$

(where r_i = age group (i) specific incidence rate). The difference in the numbers of incident cases between 2000 and 2005 was also divided into two components as in step (4): changes due to cancer risk or the population changes.

8. Results

8.1 Estimated population by area for year 2005 in China

The population pyramids by sex, age and area for 1990 and 2000, and the estimate prepared for 2005 are shown in Fig. 6. Urbanization and ageing (with 10.8% population age over 60) are the two major characteristics of the current population structure of China. (Paper III)

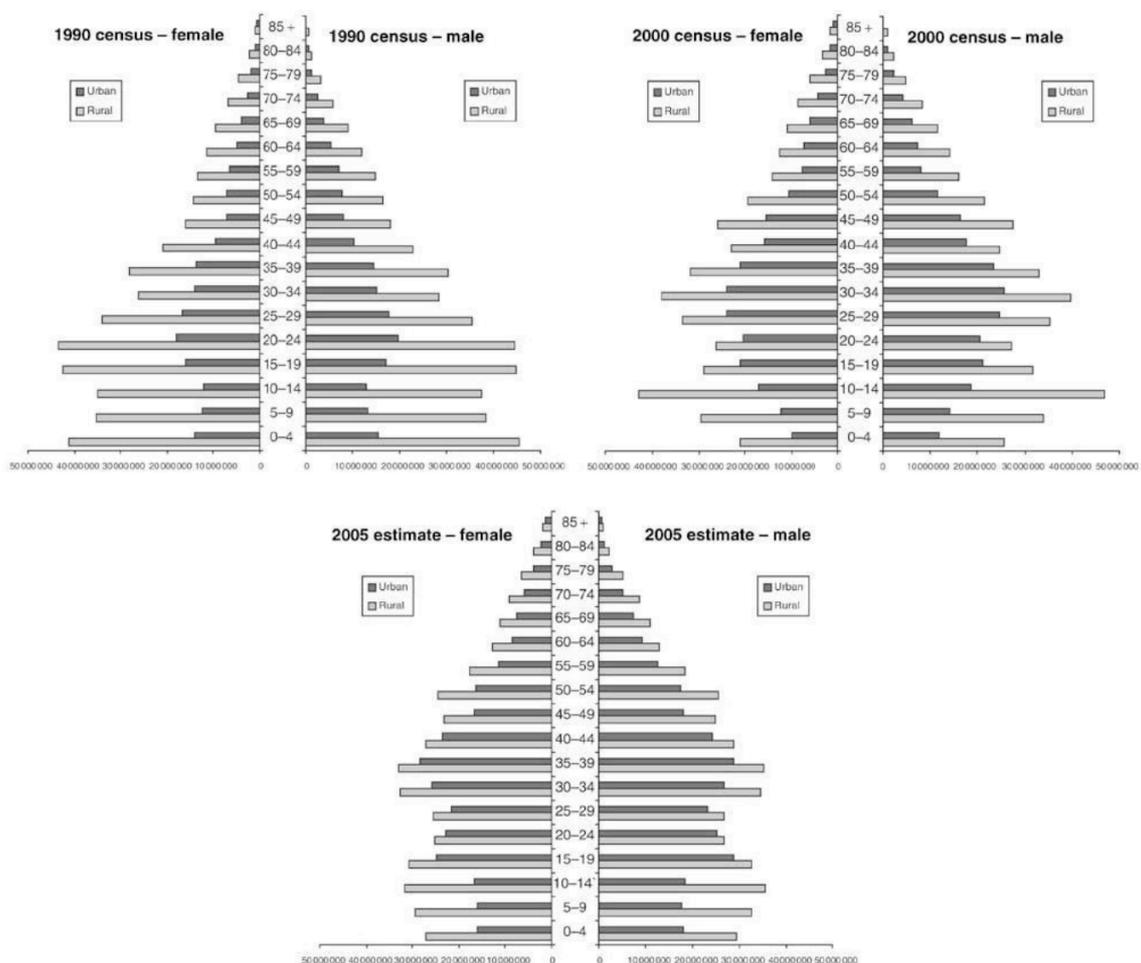


Fig. 6. Population pyramids for the years 1990, 2000, and 2005 by area, sex and age group in China.

8.2 Cancer registration practice in China

Forty-eight cancer registries currently exist in China, covering 73 million people (5.7% of the total population of China in 2000) (Paper V). The distribution of these registries is shown on the map in Fig. 7. The population size and the years of registered data for the registries with more than 15 years of incidence or mortality data held as hard copy, or more than 10 years in electronic form are shown in Table 2. The three oldest registries are in Linzhou county, Shanghai city and Qidong county. Most registries are

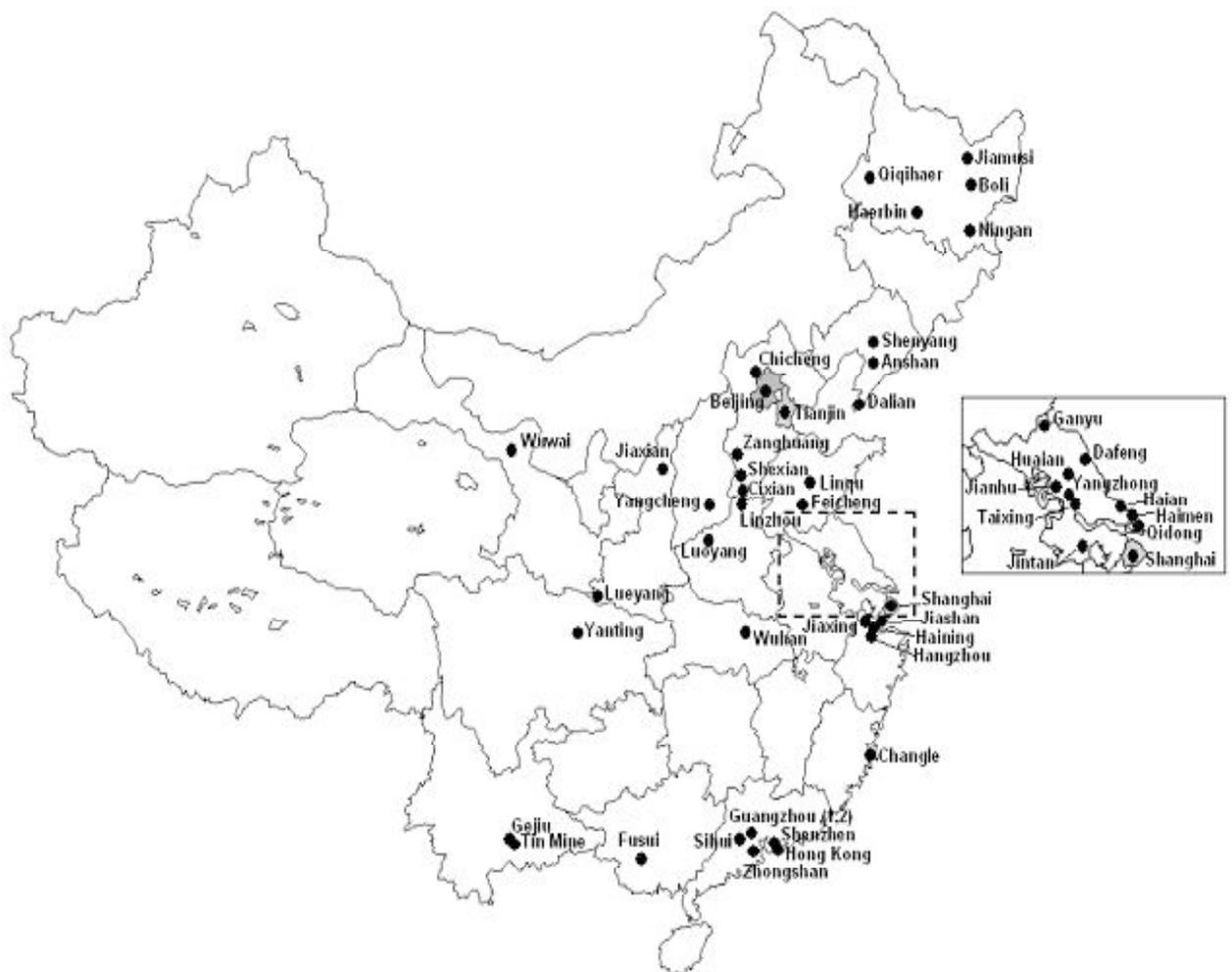


Fig. 7 Distribution of cancer registries in China, 2002

Table 2. Cancer registries with more than 15 years of incidence or mortality data held as hard copy, or more than 10 years in electronic form.

Province	Cancer Registry	Population	Inc-C (Yrs)	Inc-E (Yrs)	Mor-C (Yrs)	Mor-E (Yrs)
Beijing	Beijing	2430000	25	14	25	14
Shanghai	Shanghai	13155350	31	31	31	31
Tianjin	Tianjin	3694846	22	22	22	22
Liaoning	Dalian	5517191	16	16	21	12
	Anshan	1458963	17	10	20	10
Hubei	Wuhan	4299624	22	10	22	10
Guangdong	Zhongshan	1348300	32	32	32	32
	Sihui	396917	15	6	29	28
Jiangsu	Huaian	1232270	15	4	30	8
	Qidong	1161922	30	30	30	30
	Dafeng	740000	3	3	18	18
	Yangzhong	273177	12	12	20	10
Gansu	Wuwei	1000000	11	.	8	.
Henan	Linzhou	982254	43	43	43	43
Heilongjiang	Haerbin	884469	12	12	12	12
	Jiamusi	527854	11	0	17	0
	Boli	317151	16	0	16	0
Shangdong	Linju	786530	13	10	23	10
Fujian	Changle	680000	15	15	15	15
Zhejiang	Haining	640319	25	20	25	20
	Jiashan	380581	15	15	16	16
	Jiaxing	294999	11	11	11	2
Hebei	Cixian	625147	27	12	27	12
	Shexian	379961	10	10	20	10
	Chicheng	282051	15	0	15	0
	Zanghuang	230550	15	0	17	0
Sichuan	Yanting	608452	15	7	34	8
Yunnan	Gejiu	450000	32	20	32	20
	Yunnan Tin Mine	50000	28	10	28	10
Guangxi	Fusui	411225	17	10	17	10
Shanxi	Yangcheng	391836	3	0	30	0
Shaanxi	Jiaxian	240000	15	.	15	.
	Lueyang	201960	17	.	8	.

Note: Inc-C (Yrs): years for incidence data in hard copy; Inc-E (Yrs): years for incidence data in electronic file; Mor-C (Yrs): years for mortality data in hard copy; Mor-E (Yrs): years for mortality data in electronic file

in urban areas and located in the relatively more economically developed areas with good facilities for diagnosis and treatment, while those at county level are usually in high-risk areas for specific cancers, and supported by various cancer research programmes. Remarkable variations in practice between registries, with respect to data collection, data management and coding were found in the survey. Differences also existed in administrative aspects and sources of financial support. Both active and passive data collection methods were used in the registries. Population data were obtained mostly from the local security departments, which take responsibility for registration of vital events. Almost all registries collected the basic registration information (name, sex, date of birth, address, ethnicity, incidence date, most valid basis of diagnosis, topography, the date and cause of death). 57.0% case records were abstracted and 87.8% coded by registry personnel. 28 registries (58.3%) define ‘cancer incidence date’ as the date of first diagnosis of the cancer by a physician, 10 (20.8%) as the date of the notification report, and the rest use the date of first symptom or date of pathology report. 75% of registries collected information on all diagnostic methods, 55% registries coded the most valid basis of diagnosis. ICD-9 and ICD-10 were the most often used coding systems in registries; few registries use ICD-O to code topography. With respect to follow-up procedures, most cancer registries actively follow-up cases to ascertain vital status, and used death certificate information to obtain information on missed cases and to check or update registered data, but only half of the registries trace back death notifications before registering a case as a “death certificate only” (DCO) case. Various methods were used among registries for estimating the completeness and validity of registered data. Most registries published an annual report, as basic information for local cancer control and research activities.

8.3 Validity of cancer mortality data from available data sources in China

Neither of the ongoing data collection systems for mortality – CHIS and DSP – covers a representative sample of the Chinese population, with respect to age group, sex, and urban-rural residence (Paper II). The CHIS population has rather more individuals in the middle and old age groups (ages 30–65), and fewer children and adolescents. There is more urban population in the CHIS sample with rural/urban ratios close to 1 in all age groups, while the DSP sample has an excess of individuals living in rural areas, especially in younger and older age groups.

In terms of cancer mortality rates, using the second mortality survey as a reference, differences of 10% or more from the rates obtained in the second national

mortality survey (1990–1992) were more common with the DSP data than with the CHIS material. For most sites, the estimated national rates obtained using the CHIS data were closer to those from the national survey rates than are the DSP-based estimates (shown in Table 3) The corresponding annual percentage changes, fitted by log-linear models showed the trends obtained with the CHIS dataset to be more stable and reliable, while the DSP dataset showed some extreme trends (such as declines of 7% per year for stomach cancer in men, and falls of 16% a year for cervical cancer, both in urban areas; trends shown in Fig. 8). Furthermore, the CHIS within-stratum estimates are more stable than those of DSP, partially due to its larger sample size. Therefore, although the DSP population maybe more representative of the national population, the CHIS data is the preferred source for estimation of national cancer mortality and study of time trends, although further appropriate weighting might still be required.

Table 3. Estimated national age-standardized mortality from cancer, by sex, for year 1992, based on the datasets from the survey, DSP and CHIS in 1992 and the age standardized mortality rates from the second national mortality survey in 1990–1992 (rates for per 100 000)

Sex	Cancer	ASR-Survey	National estimates			Difference (%)	
			From CHIS	From DSP	From Survey	CHIS-Survey	DSP-Survey
Male	Nasopharynx	2.3	2.3	2.1	2.3	-0.5	-9.3
Male	Oesophagus	22.6	18.1	14.4	22.3	-19.0	-35.4
Male	Stomach	33.4	25.6	23.6	33.1	-22.6	-28.8
Male	Liver	28.7	27.8	24.0	28.6	-2.8	-16.1
Male	Lung	24.2	25.7	23.7	24.5	4.7	-3.3
Male	All	136.2	125.5	112.5	136.1	-7.8	-17.3
Male	Colon/Rectum	5.9	6.1	5.9	6.0	1.9	-0.6
Male	Leukaemia	4.0	3.7	3.8	4.0	-6.8	-3.3
Female	Nasopharynx	1.0	0.9	0.7	1.0	-10.3	-28.5
Female	Oesophagus	11.1	9.0	6.5	10.9	-17.9	-40.3
Female	Stomach	15.4	12.6	10.4	15.2	-17.5	-31.6
Female	Liver	10.3	10.0	9.4	10.2	-2.3	-8.5
Female	Lung	9.6	10.2	8.8	9.8	4.9	-9.6
Female	Breast(F)	3.3	3.5	2.9	3.3	4.3	-12.6
Female	Cervix	3.5	3.1	2.3	3.5	-12.8	-34.4
Female	All	73.0	69.2	60.3	72.9	-5.1	-17.3
Female	Colon/Rectum	4.3	4.9	4.2	4.4	12.1	-3.3
Female	Leukaemia	3.3	2.9	3.6	3.3	-10.0	9.9

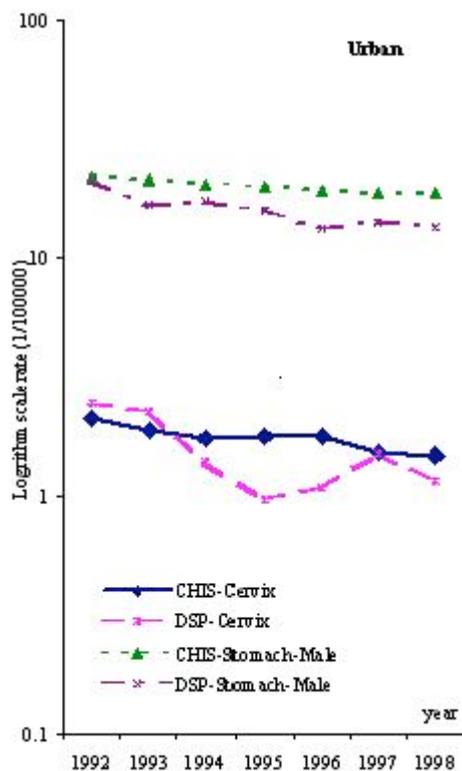


Fig. 8 The age-standardized mortality rates (per 100,000) for stomach (in male) and cervical cancer in urban areas, during year 1992 and 1998 between CHIS and DSP datasets

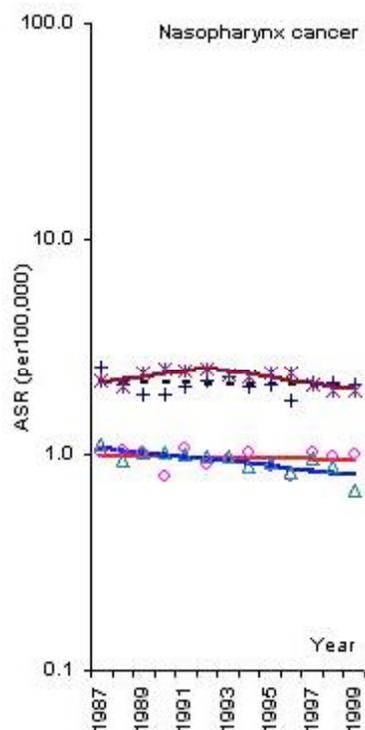
8.4 Cancer mortality and trends during 1987–1999 in China

Age-standardized mortality rates, trends and estimated annual percentage change (EAPCs) obtained by joinpoint analysis for specific common cancers were calculated by sex and area (urban and rural) during 1987–1999 (Fig. 9 and Table 4). Table 4 also shows the age-specific EAPCs from Poisson regression models. Generally, the mortality rates were higher in men than in women. Comparing rural and urban populations, many cancers had higher mortality rates in rural than in urban people, such as cancers of the oesophagus, stomach, liver and cervix uteri, while there were no major differences between the two areas for cancers of the nasopharynx and leukaemia. (Paper I)

Between 1987 and 1999 the age-standardized mortality rates for all cancers combined declined slightly in rural areas for both sexes (−0.1% per year in males and −0.4% in females), while in urban areas rates increased until 1989 (annual increase 4.6%), followed by a significant steady decrease of −1.6% until 1996 and then increased again at 1.3% per year in males; in females, mortality rates dropped significantly (−1.5% per year) until 1996 and then increased 2.2% annually. For the age-specific mortality rates there is a significant decline for age groups 0–4 and 55–64 in both sexes

and areas (Table 4). The mortality rates for cancers of the oesophagus, stomach and cervix uteri significantly decreased during the study period (except for stomach cancer before 1994 in urban males) and declined among all age groups over age 55. In lung cancer, the rates, however, dramatically increased throughout the study period in both urban and rural populations, more obviously in rural, and among all age groups in both sexes (except for males aged 55–64, females aged 45–54 and the youngest age group in urban population). Similar dramatically increasing trends were present for female breast cancer: the EAPC in rural areas was +0.9% until 1996, with a significant increase (11.6% per year) thereafter, while a steady, slight but significant increasing trend (0.9% annually) was present in urban females during the 13 years. Mortality from cancers of the colon-rectum and liver had different trends in different time periods, sub-populations and age groups (Fig. 9 and Table 4). Steadily declining trends were also present for leukaemia (except for urban males after 1996), and nasopharynx, although the mortality rates were much lower than for the above cancers (Table 4). The trends in age-specific mortality rates suggested some different trends in younger population, which may presage future overall trends – for example, increasing mortality from cancer of the cervix (for urban women aged 35–44, with a statistically significant increase of 4.1% per year).

+ Rural-Male-O ◊ Rural-Female-O * Urban-Male-O △ Urban-Female-O
 - - - Rural-Male-E — Rural-Female-E — Urban-Male-E - - - Urban-Female-E



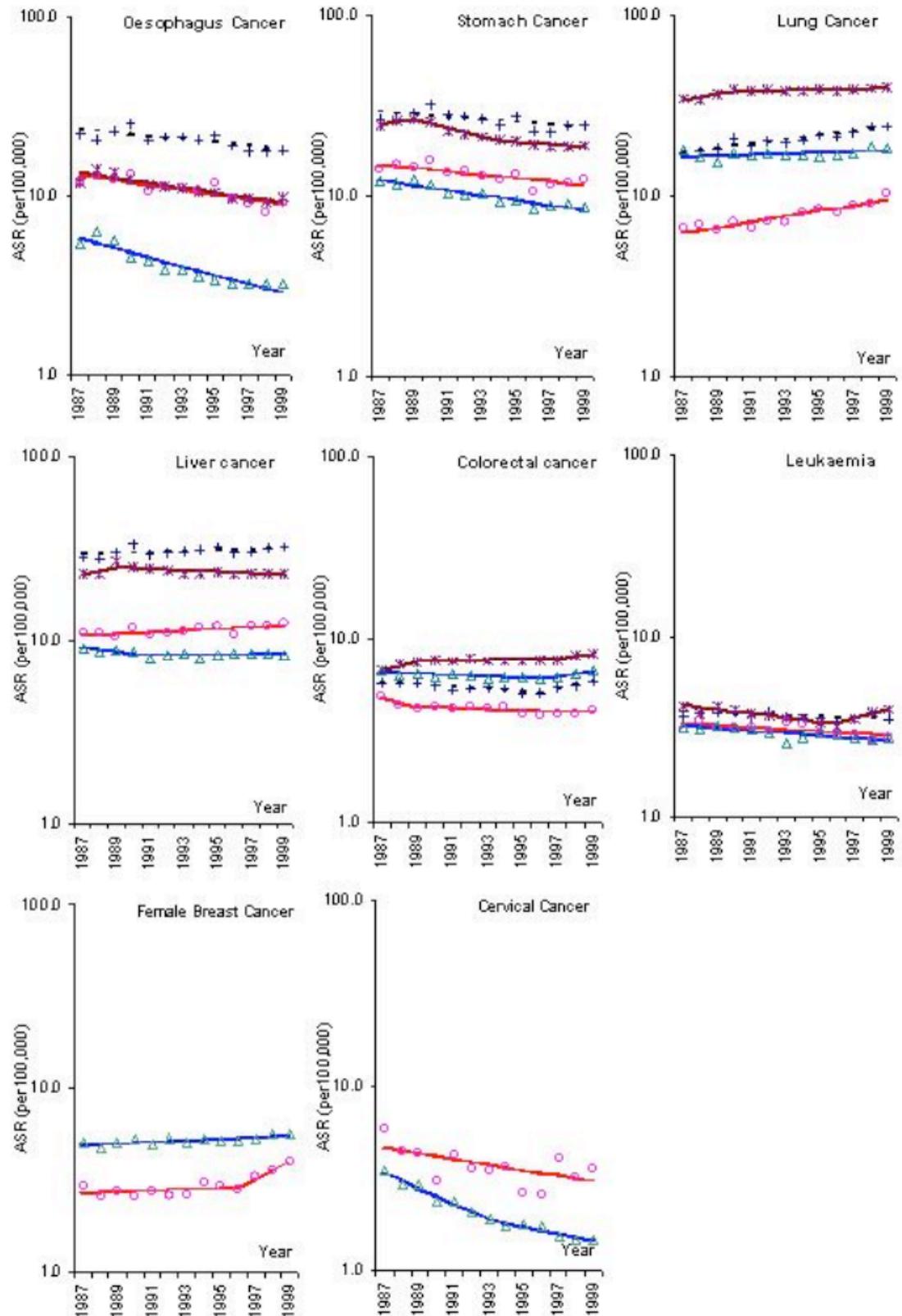


Fig. 9 Time trends for mortality rates for certain common cancers during 1987–1999 in China, using age-standardized rate (ASR) by area

Table 4. Estimated annual percentage of change (EAPC) for age-standard mortality rates by joinpoint analysis, for some common cancers by area and sex, during 1987–1999 in China

Cancer	Area-Sex	Deaths 1987	ASR 1987	Deaths 1999	ASR 1999	EAPC-all	Period-1	EAPC-1	Period-2	EAPC-2	Period-3	EAPC-3	EAPC(15-84)	EAPC(35-44)	EAPC(45-54)	EAPC(55-64)	EAPC(65-64)	
All Cancers	Rural-Male	25666	118.6	39016	123.2	-0.1	1987-1999	-0.1				-0.1	-0.8*	0.9*	0.9*	0.9*	-1.2*	
	Rural-Female	15664	69.2	22784	68.4	-0.4	1987-1999	-0.4				0.0	-0.2	0.9*	0.9*	0.9*	0.9*	-1.0*
All Cancers	Urban-Male	41262	128.6	57081	130.5	-0.6*	1987-1989	4.6	1989-1996	-1.6*	1996-1999	1.3	-2.5*	0.7	0.3	0.3	0.3	-2.4*
	Urban-Female	25181	81.5	35430	74	-0.8*	1987-1996	-1.5*	1996-1999	2.2			-2.7	0.4	-0.6	-0.6	-0.6	-1.6*
Oesophagus	Rural-Male	4788	22	5740	18.1	-2.1*	1987-1999	-2.1*				-0.5	-3.6*	-1.4*	-1.4*	-1.4*	-1.4*	-3.2*
	Rural-Female	2778	11.9	3143	9.1	-3.2*	1987-1999	-3.2*				-2.4	-6.6*	-3.3*	-3.3*	-3.3*	-3.3*	-4.0*
Oesophagus	Urban-Male	3742	12	4263	9.8	-3.2*	1987-1999	-3.2*				-5.2	-0.2	-4.6*	-0.2	-0.2	-0.2	-4.6*
	Urban-Female	1654	5.3	1609	3.2	-5.7*	1987-1999	-5.7*				-3.0	-10.2*	-7.8*	-7.8*	-7.8*	-7.8*	-7.8*
Stomach	Rural-Male	5766	26.5	7798	24.6	-1.7*	1987-1999	-1.7*				0.7	-2.0*	0.7	-2.0*	0.7	-2.0*	-2.7*
	Rural-Female	3279	14.1	4211	12.3	-2.2*	1987-1999	-2.2*				0.6	-3.1*	-1.5*	-1.5*	-1.5*	-1.5*	-2.7*
Stomach	Urban-Male	7717	24.4	8392	19.1	-3.1*	1987-1989	3.4	1989-1994	-5.4*	1994-1999	-1.4	-2.9*	-0.6	-1.6*	-1.6*	-1.6*	-4.9*
	Urban-Female	3781	12.1	4229	8.6	-3.2*	1987-1999	-3.2*				-5.4*	-0.2	-2.3*	-2.3*	-2.3*	-2.3*	-4.1*
Lung	Rural-Male	3696	17	7618	24	2.7*	1987-1999	2.7*				2.5*	2.5*	2.5*	2.5*	2.5*	2.5*	0.8*
	Rural-Female	1505	6.6	3460	10.1	3.6*	1987-1999	3.6*				1.7	1.9*	3.9*	3.9*	3.9*	3.9*	3.1*
Lung	Urban-Male	11035	34.2	17572	39.8	0.9*	1987-1990	3.8*	1990-1999	0.4		-1.4	3.1*	3.1*	0.9*	0.9*	0.9*	-1.7*
	Urban-Female	5451	17.7	8893	18.2	0.7	1987-1999	0.7				-0.6	1.1	-1.4*	-1.4*	-1.4*	-1.4*	-0.5
Liver	Rural-Male	6106	28.4	10117	31.8	0.7	1987-1999	0.7				0.3	-0.6	1.6*	1.6*	1.6*	1.6*	0.1
	Rural-Female	2377	10.9	4034	12.4	1.0*	1987-1999	1.0*				2.4*	0.6	1.6*	1.6*	1.6*	1.6*	0.1
Liver	Urban-Male	7779	23.2	10334	23.1	-0.6	1987-1989	4.7	1989-1999	-1.0*		-1.7*	0.5	0.5	0.3	0.3	0.3	-2.3*
	Urban-Female	2826	9.1	3978	8.3	-0.5	1987-1991	-2.8*	1991-1999	0.3		-1.3	-0.4	-0.4	-1.0*	-1.0*	-1.0*	-0.9
Colon-Rectum	Rural-Male	1249	5.8	1855	5.9	-0.2	1987-1996	-1.4*	1996-1999	5.0*		-1.5	-0.7	0.4	0.4	0.4	0.4	-1.0*
	Rural-Female	1119	4.8	1389	4.1	-1.1*	1987-1989	-5.8	1989-1999	-0.7		-1.5	-2.1*	0.1	0.1	0.1	0.1	-0.8
Colon-Rectum	Urban-Male	2130	6.8	3648	8.3	1.1*	1987-1989	5.5	1989-1997	0.3	1997-1999	4.0	-8.8*	0.2	2.1*	2.1*	2.1*	0.3
	Urban-Female	2124	6.8	3283	6.7	-0.1	1987-1996	-0.9*	1996-1999	3.3*		-5.6*	-1.1	0.0	0.0	0.0	0.0	-0.6
Breast(F)	Rural-Female	647	2.9	1268	4	3.1*	1987-1996	0.9	1996-1999	11.6*		4.8*	4.8*	4.8*	4.8*	4.8*	4.8*	2.9*
	Urban-Female	1608	5.1	2633	5.6	0.9*	1987-1999	0.9*				-2.2*	2.5*	2.5*	2.5*	2.5*	2.5*	1.5*
Cervix	Rural-Female	1318	5.8	1127	3.5	-3.5*	1987-1999	-3.5*				-0.8	0.1	0.4	0.4	0.4	0.4	-4.3*
	Urban-Female	1086	3.5	703	1.5	-7.1*	1987-1993	-9.4*	1993-1999	-4.3*		0.3	4.1*	4.1*	4.1*	4.1*	4.1*	-9.5*
Nasopharynx	Rural-Male	548	2.5	665	2.1	-0.5	1987-1999	-0.5				-0.9	-1.2	-0.5	-0.5	-0.5	-0.5	-0.2
	Rural-Female	220	1	322	1	-0.3	1987-1999	-0.3				-5.6*	-0.7	3.2*	3.2*	3.2*	3.2*	-1.0
Nasopharynx	Urban-Male	740	2.2	881	2	-1	1987-1992	3.2	1992-1999	-3.4*		-2.9*	-1.5	0.3	0.3	0.3	0.3	-1.9
	Urban-Female	347	1.1	318	0.7	-2.4*	1987-1999	-2.4*				-4.1*	-2.0	-1.8*	-1.8*	-1.8*	-1.8*	-1.8
Leukemia	Rural-Male	738	3.6	1002	3.4	-0.7*	1987-1999	-0.7*				-1.1*	-1.1	-1.1	-1.1	-1.1	-1.1	-1.0
	Rural-Female	617	3	778	2.7	-1.3*	1987-1999	-1.3*				-1.3	-0.9	0.5	0.5	0.5	0.5	-1.8*
Leukemia	Urban-Male	1225	4.1	1434	3.9	-0.9	1987-1996	-2.6*	1996-1999	6.7*		-1.9*	-1.6*	-0.1	-0.1	-0.1	-0.1	-2.0*
	Urban-Female	888	3.1	1002	2.7	-1.5*	1987-1999	-1.5*				-3.0*	-1.5	0.7	0.7	0.7	0.7	-1.7

ASR: age-standardized mortality rates adjusted to the 1990 China census population (per 100,000); + EAPC: estimated annual percentage of change; * The EAPC differs significantly from 0 (P<0.05)

8.5 Cancer mortality and incidence estimates and projection for 2000 and 2005

8.5.1 Sensitivity analysis

The results of the sensitivity analysis of the methods used for estimating the mortality for four cancers using CHIS data in 1999 are shown in Table 5, comparing the observed and estimated number of deaths and age-standard mortality rates (with 95% estimated intervals) produced by different models. The mortality rates estimated for 1999 by the log-linear model (which assumes different proportional changes for different age groups) were much closer to the observed values with the narrowest confidence intervals, both for numbers of deaths and for the age-standardised rates, compared with the other three prediction models (Table 5). The non-fitted values for lung and breast cancer from the simple non-linear model (the 4th model shown in sensitivity analysis) may be due to iterations not leading to likelihood maximum. (Paper III)

According to the sensitivity analysis for the cancer incidence prediction (shown in Table 6), the national incidence estimates based on the ratios of incidence to mortality from seven aggregated registries data (Pooled (7)) were slightly greater than those based on the full stratified model (Urban + Rural(4)). The estimates based on a stratified analysis with only Jiashan and Qidong contributing to the rural model (Rural (2)), however, was very different, Even the aggregated model with only these two registries data contributing to the rural area (Pooled(5)) provided results closer to the full stratified model (Urban + Rural(4)) than did the Urban + Rural(2) model based on the stratified analysis. (Paper IV)

Table 5. Observed and estimated number of deaths, age-standard mortality rates with 95% estimated intervals produced by different models for four cancers in CHIS data for 1999

Cancer	Model	Numbers of Deaths				Age-standardized mortality rates (1/100,000)			
		Observed	Estimated	95% Confidence Interval		Observed	Estimated	95% Confidence Interval	
Lung	$\log E[M_{it}] = \alpha_i + \beta_i t$	3460	3218.8	3067.7	3369.8	12.6	11.7	11.2	12.3
	$\log E[M_{it}] = \alpha_i + \beta t$	3460	3217.6	3066.7	3368.5	12.6	11.8	11.2	12.3
	$E[M_{it}] = \alpha_i + \beta_i t$	3460	3161.7	2906.1	3417.4	12.6	11.5	10.6	12.5
	$E[M_{it}] = \alpha_i(1 + \beta t)$	3460	2605.3	2473.4	2777.2	12.6	9.6	9	10.1
Breast	$\log E[M_{it}] = \alpha_i + \beta_i t$	1268	1127.6	1033.6	1221.6	4.8	4.3	3.9	4.7
	$\log E[M_{it}] = \alpha_i + \beta t$	1268	1121.9	1013.5	1230.4	4.8	4.2	3.8	4.7
	$E[M_{it}] = \alpha_i + \beta_i t$	1268	1104.6	957.4	1251.8	4.8	4.2	3.7	4.8
	$E[M_{it}] = \alpha_i(1 + \beta t)$	1268	946.3	867.1	1025.4	4.8	3.6	3.3	3.9
Oesophagus	$\log E[M_{it}] = \alpha_i + \beta_i t$	1690	1442.1	1328.1	1556.1	4.1	3.7	3.4	4
	$\log E[M_{it}] = \alpha_i + \beta t$	1690	1434.9	1313.7	1556	4.1	3.7	3.4	4
	$E[M_{it}] = \alpha_i + \beta_i t$	1690	1420.2	1051.6	1788.9	4.1	3.6	2.7	4.6
	$E[M_{it}] = \alpha_i(1 + \beta t)$	1690	1413.6	1276.5	1550.8	4.1	3.6	3.3	4
Cervix	$\log E[M_{it}] = \alpha_i + \beta_i t$	703	657	595.2	718.8	1.9	1.8	1.6	1.9
	$\log E[M_{it}] = \alpha_i + \beta t$	703	642.4	550.5	734.3	1.9	1.7	1.4	1.9
	$E[M_{it}] = \alpha_i + \beta_i t$	703	614.9	510.6	719.2	1.9	1.7	1.4	1.9
	$E[M_{it}] = \alpha_i(1 + \beta t)$	703	619.9	514	725.7	1.9	1.6	1.3	1.9

8.5.2 Estimation and projection of cancer patterns in 2000 and 2005

The estimated and projected number of deaths and cases for 10 common cancers and all other cancer combined in China for the years 2000 and 2005 are summarized in Figs. 10 and 11. The corresponding age-specific (above age 44) and age-standardized rates adjusted by the world population, are shown in Table 7.

Compared with age-standardized mortality rates in 1991, based on the national mortality survey in 1990–1992, it was predicted that mortality rates for all cancers combined would change little between 1991 and 2005 (–0.8% in men and +2.5% in women), but population growth and ageing would result in an increasing absolute number of deaths. The total numbers of deaths from all cancers will increase from 1.2 million in 1991 (752,000 in males, 422,000 in females) to 1.5 million in 2000 (993,000 in males, 550,000 in females), and 1.8 million in 2005 (1.1 million in males and 649,000 in females), the largest increase is in female breast cancer (from 19000 deaths in 1991 to 48,000 in 2005, an increase of +155.4%), followed by lung cancer (+112.1% in men, +153.5% in women), and liver cancer (+58.4% in men, +72.2% in women) (Fig 10). Increases in numbers were observed in almost all age groups and the dramatic increases in the rates in populations younger than age 55 imply that there will further increases in risk from these cancers in the future (Table 7). (Paper III)

Table 6. Sensitivity analysis for the cancer incidence predictions

Cancer	Sex	Incidence/mortality model						
		Urban	Rural (2)	Rural (4)	Urban + rural (2)	Urban + rural (4)	Pooled (5)	Pooled (7)
Estimated age-standardized (ASR world) incidence per 100,000								
Esophagus	Male	15.6	28.3	33.8	23.9	27.5	25.9	27.6
Stomach	Male	31.5	43.1	45.3	39.0	40.5	42.3	41.9
Esophagus	Female	5.3	13.1	15.7	10.4	12.0	11.3	12.1
Stomach	Female	15.2	19.9	20.9	18.2	18.9	19.7	19.5
Estimated number of cancer cases (thousands)								
Esophagus	Male	32.9	112.5	135.0	145.4	167.9	157.4	168.5
Stomach	Male	66.7	172.0	181.0	238.7	247.7	358.9	256.3
Esophagus	Female	11.5	54.0	63.6	65.5	75.1	71.1	75.3
Stomach	Female	33.2	80.4	84.5	113.7	117.7	122.7	121.5

NOTE: Incidence estimates for esophagus and stomach cancers (China 2000), based on three different models of mortality/incidence: Urban, data from Beijing, Shanghai, and Wuhan cancer registries; Rural (2), data from Jiashan and Qidong cancer registries; Rural (4), data from Jiashan, Qidong, Cixian, and Changle cancer registries; Pooled (5), pooled data from Beijing, Shanghai, Wuhan, Jiashan, and Qidong cancer registries; Pooled (7), pooled data from Beijing, Shanghai, Wuhan, Jiashan, Qidong, Cixian, and Changle cancer registries.

A total of 2.1 million incident cancer cases were estimated for the year 2000 (1.3 million in men, 0.8 million in women) and these numbers were expected to increase by 14.6% by 2005. Only cancers of the oesophagus (for both sexes) and stomach (in men) showed declines in numbers of cases during these five years (Fig. 11). The five leading cancers in the year 2000, in terms of incidence, were cancers of the lung, stomach, liver, oesophagus and colon-rectum for men, and cancers of the breast, stomach, lung, liver and oesophagus for women. During 2000 and 2005, increasing incidence rates are anticipated for cancers of the lung, liver, colon-rectum, prostate and leukaemia in men, and for breast, lung, liver, colon-rectum, and cervix in women while decreases are shown for cancers of stomach, oesophagus and nasopharynx in men, and for oesophagus, stomach, nasopharynx and leukaemia in women. In addition, the increasing incidence rates of lung cancer (in both sexes) and female breast cancer indicated there would be much greater increases in the numbers of cases at these two sites (26.9% for lung cancer in men, 38.4% for lung and breast cancer in women (+38.5%)) (Fig. 11). According to the estimates, 120,000 more new lung cancer cases would occur in 2005 than in 2000 in China (0.38 million in 2000 and 0.50 million cases in 2005), and the total number of lung cancer cases would increase 26.9% in men and 28.4% in women (Fig. 11). Overall cancer incidence rates (all sites combined) were predicted to increase slightly from 2000 to 2005 in both sexes, from 209.2 to 210.8 per 100,000 in men and from 133.6 to 140.6 per 100,000 in women. The increases were more significant in the older age groups (over age 65) (Table 7). (Paper IV)

For both incidence and mortality, the largest increases are in cancers of female breast and lung cancer for both sexes. For female breast cancer, with increasing trends among all age groups, more marked increases are shown for age group 45–64 which was responsible for a 27.5% increase in the number of incident cases for that disease, while population growth and ageing are expected to contribute a further 11% increase. For lung cancer, increases are also shown among all age groups (except for the age group 55–64) in both sexes, the oldest age group (over age 75) has the highest rates for both incidence and mortality.

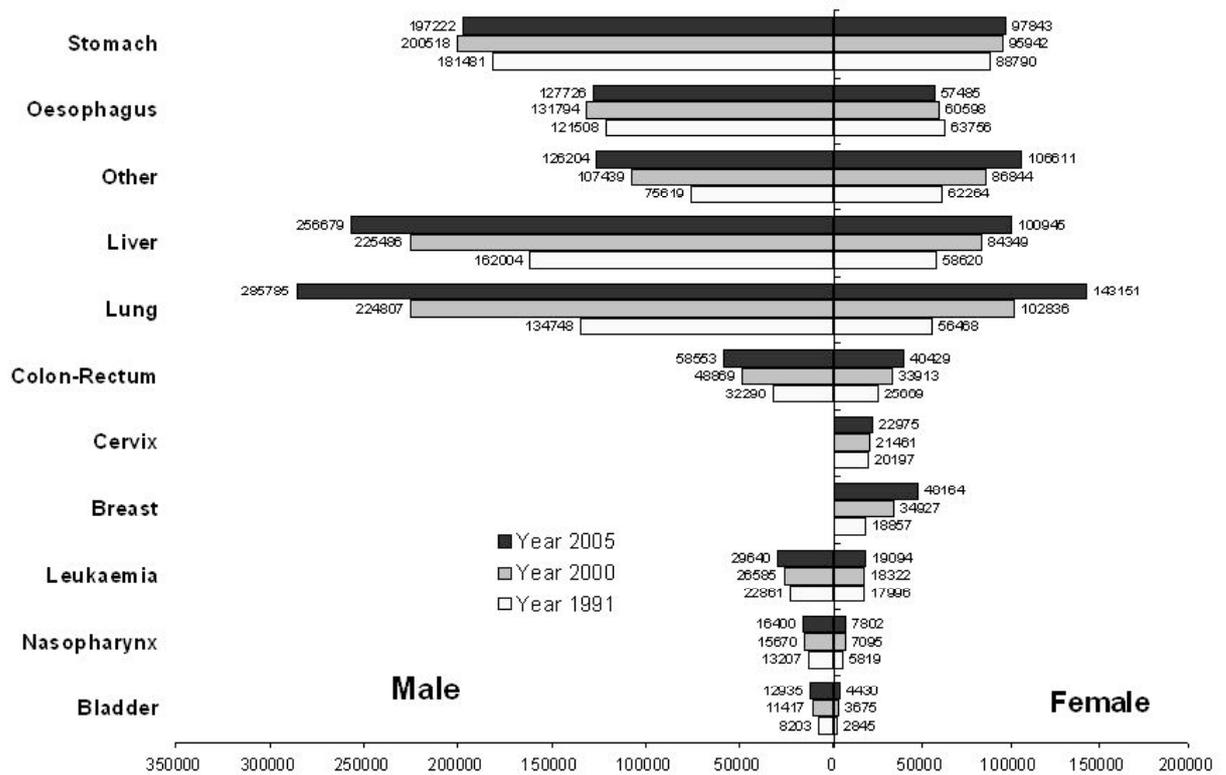


Fig. 10. Estimated and projected numbers of deaths from some common cancers in 1991, 2000 and 2005, by sex, in China

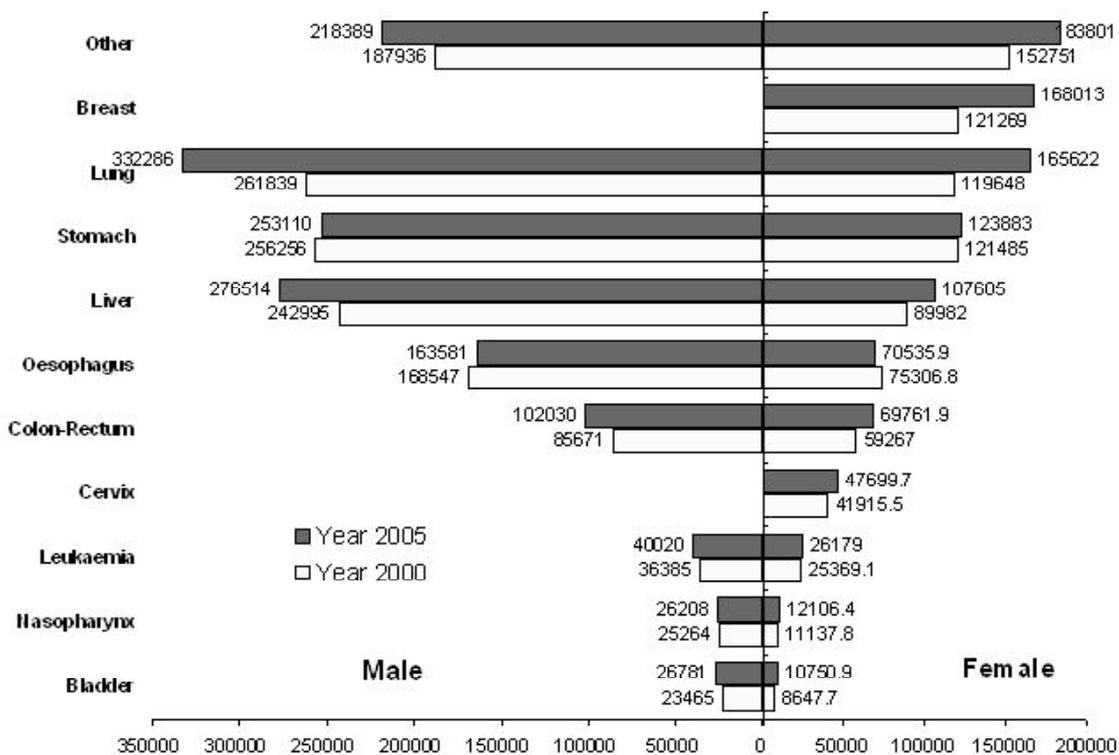


Fig. 11. Estimated and projected numbers of new cases of common cancers in year 2000 and 2005 by sex in China

Table 7. Cancer incidence and mortality rates in China for year 2000 and 2005, by sex

Cancer site	Sex	Mortality					Incidence				
		45-54	55-64	65-74	75+	ASR	45-54	55-64	65-74	75+	ASR
Year 2000											
Stomach	Female	20.5	46.6	104.6	133.3	15.0	31.6	62.1	120.5	133.0	19.5
Oesophagus	Female	11.0	32.1	69.1	99.1	9.5	18.0	44.1	79.5	95.3	12.1
Lung	Female	24.8	52.1	107.7	142.4	16.1	32.6	62.7	118.9	144.0	19.1
Liver	Female	24.5	43.6	73.8	85.8	13.3	27.1	46.6	76.3	85.8	14.5
Colorectal	Female	7.1	13.9	34.8	52.8	5.2	15.6	25.6	53.9	68.8	9.5
Breast	Female	15.2	17.9	18.3	20.2	5.5	59.6	50.6	37.3	29.7	19.9
Cervix	Female	10.0	9.4	14.0	15.4	3.4	22.8	16.1	17.9	14.6	6.8
Bladder	Female	0.7	1.5	3.3	7.5	0.6	2.6	4.1	6.0	9.5	1.4
Nasopharynx	Female	2.7	3.2	5.4	5.7	1.1	4.9	4.6	6.3	5.3	1.8
Leukaemia	Female	4.1	4.6	5.8	6.1	3.0	5.6	5.7	6.5	6.2	4.2
All sites	Female	141.4	263.0	515.8	673.9	86.3	265.6	388.3	632.6	709.4	133.6
Stomach	Male	45.5	109.4	229.4	289.6	32.7	70.0	145.7	264.3	288.8	41.9
Oesophagus	Male	27.0	66.9	167.8	210.0	21.6	44.4	92.0	193.0	202.0	27.6
Lung	Male	46.8	108.5	277.7	347.0	36.7	61.5	130.6	306.4	350.9	43.0
Liver	Male	79.5	108.3	175.2	187.8	35.3	88.0	115.8	181.2	187.8	38.9
Colorectal	Male	9.0	19.9	56.3	90.3	7.9	19.7	36.7	87.1	117.7	14.0
Bladder	Male	1.3	3.9	14.0	34.4	1.9	5.0	10.2	25.6	43.6	3.9
Nasopharynx	Male	5.7	7.8	11.4	11.3	2.4	10.4	11.4	13.2	10.5	4.0
Leukaemia	Male	5.0	6.7	11.5	13.8	4.2	6.8	8.2	12.9	14.0	5.7
All sites	Male	243.0	477.0	1048.9	1353.9	160.2	355.9	629.6	1230.3	1403.5	209.2
Year 2005											
Stomach	Female	19.0	40.1	91.3	116.3	13.3	29.3	53.4	105.2	116.0	17.4
Oesophagus	Female	8.5	25.6	55.7	86.6	7.7	13.9	35.2	64.1	83.3	9.7
Lung	Female	30.9	58.6	128.9	183.4	19.3	40.6	70.5	142.3	185.5	22.9
Liver	Female	26.2	44.5	76.4	89.4	13.9	29.0	47.6	79.0	89.4	15.3
Colorectal	Female	7.7	13.6	37.3	55.3	5.4	16.8	25.0	57.7	72.0	9.7
Breast	Female	20.1	21.8	20.8	20.5	6.7	78.7	61.4	42.4	30.2	24.8
Cervix	Female	11.3	7.5	9.7	10.9	3.2	25.8	12.9	12.4	10.3	7.0
Bladder	Female	0.8	1.7	2.9	7.6	0.6	3.2	4.4	5.4	9.6	1.5
Nasopharynx	Female	2.8	2.8	5.4	5.4	1.1	5.2	4.1	6.3	5.1	1.8
Leukaemia	Female	4.2	4.3	5.7	6.7	2.8	5.7	5.3	6.4	6.8	4.0
All sites	Female	154.0	257.8	522.4	701.0	88.8	296.9	384.7	643.5	740.1	140.6
Stomach	Male	43.3	89.4	198.3	261.9	28.8	66.7	119.0	228.5	261.2	37.1
Oesophagus	Male	25.0	52.0	147.8	188.5	18.8	41.1	71.4	170.0	181.3	24.0
Lung	Male	55.7	106.5	321.8	430.2	41.8	73.2	128.2	355.0	435.0	49.0

Liver	Male	84.3	101.2	183.0	200.9	36.2	93.2	108.2	189.3	200.9	40.0
Colorectal	Male	10.2	19.7	63.1	98.6	8.6	22.3	36.1	97.6	128.4	15.0
Bladder	Male	1.5	3.6	14.1	35.1	1.9	5.7	9.6	25.8	44.4	4.0
Nasopharynx	Male	5.5	6.8	11.4	11.2	2.3	10.0	9.9	13.3	10.4	3.8
Leukaemia	Male	5.5	6.7	13.3	17.8	4.4	7.5	8.2	14.9	18.0	5.9
<u>All sites</u>	Male	256.6	430.4	1066.4	1433.8	161.1	375.1	568.1	1252.0	1490.1	210.8

* ASR: Age Standardized Rate using world standard population as a reference; rates for per 100,000

9. Discussion

Although there have been earlier studies of cancer patterns in China, most of them have focused on specific cancer sites or on the populations of certain regions (Gao et al. 1981, Lu et al. 1985, Lu et al. 1988, Jin et al. 1999, He et al. 2001). Using information from a national mortality survey, a routine mortality surveillance system and cancer registries, and fitting different statistical models, national mortality trends for the years 1987-1999, and estimation and projection of both incidence of and mortality from cancer in China for year 2000 and 2005 at the national level were done in this thesis. It demonstrates the current cancer burden in China, thereby providing directions for cancer research, while the short-term predictions of cancer statistics, based on past trends, and a useful baseline or target for evaluating the success or failure of cancer control interventions in the country. Nevertheless, several considerations should be taken carefully into account when interpreting the results, in terms of data sources, methodological and different factors (both demographic and cancer risk factors) influencing the cancer occurrence in China.

9.1 Considerations of information sources

When no national incidence and mortality statistics are available in a region or country, as in China, the surveillance of cancer patterns from a representative sample of the population can provide a second best solution. The global cancer burden estimates, developed by IARC (Parkin et al. 1999, Pisani et al. 1999, Ferlay et al. 2001, Parkin 2001) and the GBD studies of WHO (Murray et al. 1994, Murray and Lopez 1996, Murray and Lopez 1997, Mathers et al. 2002, Shibuya et al. 2002, WHO 2005e), both represent previous attempts to explore the national cancer incidence and mortality profile in China. When estimating the worldwide incidence and mortality of 25 major cancers in 1990, Parkin et al. (1999) used the DSP dataset for the cancer mortality in China, and obtained incidence rates by using relative survival information derived from local cancer registries, as published by Sankaranarayanan et al. in 1998 (Parkin et al. 1999, Pisani et al. 1999). In GLOBOCAN 2000, CHIS data from 1998, by age, sex and area, was used as national 'representative' mortality data, and the incidence to mortality ratio method was used to estimate the incidence of cancer in China. The cancer

registries included in the model were Beijing, Cixian, Qidong, Shanghai and Wuhan. The final estimates (both for incidence and mortality) were then calculated by weighting the age, sex and site-specific rates in urban and rural areas according to the urban: rural population distribution in DSP population sample of 1997. (Ferlay et al. 2001, Parkin 2001). In the GBD 1990 study, DSP data for year 1991 was used for preliminary estimates of disease burden in China. To allow for under-registration and population distribution, a correction was applied, using the following equation:

$$D_{akj}^N = D_{aki}^N \left(\frac{1.10D_{akj}^{DSPURBAN} + 1.15D_{akj}^{DSPRURAL}}{1.10D_{aki}^{DSPURBAN} + 1.15D_{aki}^{DSPRURAL}} \right)$$

where D referred to the number of deaths, a to age group, k to sex, j for special cause studied, i for all causes, N to national mortality estimates, $DSPURBAN$ and $DSPRURAL$ referred to the urban and rural data in DSP system respectively. (Murray et al. 1994, Murray and Lopez 1996, Murray and Lopez 1997) In the GBD 2000 study, both DSP and CHIS datasets were used to evaluate the cancer patterns of mortality, GLOBOCAN 2000 was the source of the cancer incidence information in China. (Mathers et al. 2002, Shibuyal et al. 2002, WHO 2005e)

Both estimates used the routine reported mortality data sources available in China – the CHIS and DSP datasets. However, neither of these is entirely satisfactory as a representative sample for estimating national cancer patterns, as shown by the comparison in this study, and also in a recent review according to a set of criteria for evaluating the quality of national mortality statistics. (Rao et al. 2005) Although the DSP population is more representative of the national population distribution, the CHIS dataset has the advantage over the DSP data of apparently more stable cancer mortality changes (with fewer large and implausible changes), and a much larger sample size ensuring precision of estimated rates within the strata. Better estimation of cancer mortality rates at the national level requires adjustment to take account of the biased age and urban/rural composition.

Neither of the GLOBOCAN and GBD projects made use of the mortality information from the second national mortality survey in China, although this represents the most recent representative national sample of deaths. The new estimates of mortality on national level described in this thesis, however, are based on the second national mortality survey in 1990–1992, following the time trends in cancer mortality derived from the CHIS dataset. In other words, only time trends from the sample registration system – CHIS data – were used for the national mortality estimation. The less representative CHIS data would not practically influence the results. From our comparison and also a recently published paper (Rao et al. 2005), the mortality rates

from CHIS data have been considered consistent over time, and are more stable than DSP dataset due to the larger sample size.

In order to estimate incidence, the cancer registry data used in this thesis were derived from seven population-based registries for which results were published in the 8th edition of CI5 (Parkin et al. 2002), rather than the five registries involved in the GLOBOCAN 2000 estimates. Three of them are located in large, modern cities (Beijing, Shanghai, Wuhan) while the four others are from rural areas with high incidence rates for certain cancers (Qidong for liver cancer, Jiashan for colorectal cancer, Changle and Cixian for oesophagus and stomach cancers). These seven cancer registries take account of only a very small and, obviously, non-representative sample of the national population in China. Therefore, they are quite inadequate to evaluate national patterns of cancer incidence and survival. Instead, they are used solely for the purpose of estimating the ratio of mortality to incidence (by site, sex and age group) at the national level. Although the results were weighted when estimating the national cancer incidence to reduce the influence of larger urban registries on the pooled incidence/mortality ratios, it seems possible that the health care system and health facilities in these areas are likely to be superior to the national average; the estimation may therefore overestimate the national cancer incidence profile. Besides the above registries, more registries currently exist in China, as shown in the first survey for cancer registration practice. Unfortunately, their data were at present not ready to be used to estimate the national cancer pattern, either due to their being newly established or because of the poor data quality for various reasons as shown in the survey.

The information on population size and age structure was derived from the national census data, which, however, gives results by sex, age and area only for 1990 and 2000, while the population in 2005 was an estimate prepared by the UN Population Division (2003) and not broken down by urban/rural areas. Comparing the estimated 2003 UN population with the 2000 census data, some big discrepancies in the childhood age ranges were found, for example, there was 27 million more children aged 5–9 in 2005 than children aged 0–4 in 2000, presumably due to erroneous projections of fertility. Nevertheless, this bias would have only very little effect on cancer projections, since the great majority of cancers occur in the older age groups.

The national mortality and incidence estimates described in this thesis were published in the accompanying papers (Paper III and Paper IV) and were used by IARC in its new version of GLOBOCAN 2002 (Ferlay et al. 2004).

9.2 Methodological considerations

The reliability of estimation or prediction of disease occurrence depends directly on the choice of model and the explanatory variables fitted. Several statistical models were tested in this thesis. For the time trends analysis, only 13 years' cancer mortality data were available for study, so that separate period and cohort effects could not be identified. The joinpoint model, however, can provide a useful summary of the direction and size of the trend in recent years and detect when a significant change in trend occurs and has, therefore, been used to analyse cancer mortality trends.

The Poisson assumption for the number of cancer cases has been shown to be superior to methods that assume a normal distribution for the rates, experience from Finnish study (Dyba and Hakulinen, 2000). Therefore, the prediction models we chose in the thesis specified Poisson errors for the age, sex and period-specific numbers of deaths or incident cases, giving the assumption that the observed rates of change within age groups would be continued into the near future. However, as discussed by Jensen et al. (1990), considerable extra-Poisson variation may exist, which makes the estimation conservative. In the case of the predictions in this thesis, as shown in previous studies in Europe (Jensen et al. 1990, Black et al. 1997, Bray et al. 2002), overdispersion may have no practical influence on the validity of the average relation obtained and the extra-Poisson variations do not bias the regression coefficients.

In terms of the prediction period for cancer incidence and mortality, extrapolations of past trends beyond a decade or two are likely to be quite meaningless. For China, in particular, the huge changes in the socio-economic situation, diet, lifestyle and environmental factors mean that all but the most short-term predictions based upon past trends are likely to be unrealistic. Therefore, it would not be successful for a long-term projection in China, especially if there were strong cohort effects present in a disease. Nevertheless, even with the short projected period (6 years, from 1999 to 2005) in this thesis, some changes that were unexpected and difficult to explain were found: for example, there was an implausible declining mortality trend of colorectal cancer in rural females, and a much more rapidly increasing mortality trend for breast cancer in rural than in urban females.

Various models have been developed for predicting future cancer patterns based upon past trends (Teppo et al. 1974, Osmond 1985, Clayton and Schiffers 1987a, Clayton and Schiffers 1987b, Hakulinen and Dyba 1994, Hakulinen 1996, Dyba and Hakulinen 2000). To replicate as closely as possible the changes observed in the CHIS mortality data set 1991–1999, the sensitivity analysis showed that a log-linear model with age-specific proportion changes gave results closest to the observed values and

with the narrowest confidence intervals. Therefore it was used to predict national mortality in 2000 and 2005 in the thesis. Although a markedly increasing trend might lead to an ‘explosive’ forecast when using a logarithm scale (Teppo et al. 1974), with the short 6 years projection period in our study (1999–2005) this probability would be quite small, and in fact the results show that the final age-specific patterns of mortality appeared to be reasonable for the most part. Nevertheless, some strange patterns still emerged. For instance, the mortality rates for cervical cancer in 2005 would be highest in the age group 45–54 (11.3 per 10⁵), decrease at ages 55–64 (to 7.5 per 10⁵), and then rise again (to 10.9 per 10⁵ at ages 75+). This was the consequence of large rates of change in the CHIS data between 1991 and 1999 that differ by age. Large increases in mortality rates for cervical cancer in younger women quite likely indicate a set of birth-cohort specific changes in the risk of this disease.

With only sparse information on population-based survival in China, it is impossible to estimate incidence from mortality and survival (Verdecchia et al. 2001). National mortality estimates combined with the aggregates of cancer registries’ incidence and mortality data were used instead, to achieve stable incidence estimates at the national level. This method has widely been used, and is considered accurate for the purpose (Jensen 1990, Parkin et al. 1999, ENCR 2001, Bray et al. 2002). Nevertheless, the non-representative nature of the registries data needs to be taken into account. Technically, it should be reduced by stratifying urban and rural populations estimates separately. The sensitivity analysis proposed in the study for two sites – oesophagus and stomach – suggested that a pooled model (with 3 urban and 4 rural cancer registries) did indeed slightly (0.5–3%) overestimate incidence. However, for most sites, the small sample size of the data from the only two rural cancer registries (Jiashan and Qidong) was no more likely to be representative of rural populations than the pooled (crude) analysis. A pooled set of ratios from all cancer registries was, therefore, chosen to estimate incidence from national mortality, as being more reasonable and having higher validity than the stratified analysis. In addition, technically, the systematic errors from the method itself needed to be considered, as discussed in earlier publications (Jensen 1990, Bray et al. 2002). Ideally, if reporting of cause of death were completely accurate, the ratio of mortality to incidence would equal (1-survival probability) in a steady state, with constant incidence and survival. As indicated before, cancer incidence rates and survival are, however, changing in China, and furthermore, no information on the validity of the cause of death is available for the routine mortality data used in the study. If there were systematic differences in classifying the causes of death between the areas covered by the cancer registries included in the model “equation area”, and the national-level (“estimation area”), the predictions would be distorted. The equation area in the thesis was based on data from seven cancer registries in which three urban registries all

located in large, modern conurbations (Beijing, Shanghai, Wuhan), with probably superior health care systems and health facilities than the national average level “estimation area”. As a result, the method might overestimate the national cancer incidence profile.

9.3 Determinants of changing cancer burden in China

Projected values come true to the extent that they are not modified by a variety of factors that affect the risk of cancer, or its outcome, in the population concerned. Growing knowledge concerning systematic and dynamic variations in cancer mortality and incidence according to environmental exposures, socio-economic status, life-style, health care interventions and other factors, not only offers the opportunity for setting goals and targets in health questions about cancer, but can also potentially improve the accuracy and reliability of estimations and projections of disease statistics. Theoretically, projections of cancer incidence or mortality rates based on past trends are not necessarily meant to predict what will actually happen in the future (Hakulinen 1996). With a dramatically improved economic and health care situation (with respect to primary health care, health services and education, public hygiene, medical diagnostic methods and treatment techniques, the quality of registration and classification of the causes of death) during the last three decades, profound changes are occurring with respect to environment and lifestyle in China. These will result in dynamic changes of the cancer pattern in China, both in incidence and mortality.

The difference between urban and rural areas shown in the patterns and time trends of cancer mortality, such as urban populations with significantly higher mortality rates for lung, colon-rectum and female breast cancers than rural populations, and lower rates for oesophageal, stomach, liver and cervix cancers, provide clues to the likely importance of differences in socio-economic circumstances, lifestyle, diet and health service provision between two areas, as well as the substantial migration of populations within China. According to official statistics, the percentage of the population living in urban areas increased from 19.4% in 1980 to 30.4% in 1998 (SSBC 1999) and was estimated to be 36.2% in 2000 (UN-SD 2002). Three factors were concluded by Heilig (1999) as driving the increasing urbanization in China: the huge “excess population” in agriculture, the income gap between rural and urban employment, and the growing labour demand of urban industries and service sectors. Furthermore, it should be noted that besides the above cited official statistics, there is a large number of temporary laborers and a certain amount of “illegal” rural-urban migration that is not registered or may have been tolerated; for example, Beijing and Shanghai each have a “floating”

population of 2–3 million people (Heilig 1999). This substantial migration of populations within China is likely to result in significant changes in cancer mortality and incidence patterns within rural and urban areas.

The remarkable changes in mortality and incidence during recent decades in China involve most of the major cancers, and the reasons for the changes in rates are largely based on the dramatic changes in the socio-economic, environmental, lifestyle, nutrition, and health care situation following the economic reforms started two decades ago. During 1990 and 2000, the per capita gross domestic product increased from US\$339 to US\$853, the percentage of GDP devoted to health care increased from 4.1% to 5.3% and per capita health care costs increased from 76.6 to 376.4 Chinese Yuan (UN-SD 2005, CHIS-MOH 2005). Along with the improving socio-economic status, marked changes in the organisation and provision of health care facilities and human resources also occurred in China during the last decade, which allowed more extensive and effective early detection, diagnosis, and treatment for cancer, and thus may have contributed to decreasing mortality and improved prognosis (Sankaranarayanan et al. 1998).

The very large improvements in socio-economic circumstances, educational level, sanitary conditions of housing and daily life may be responsible for declines in cancers associated with infection such as non-cardia gastric cancer related to *Helicobacter pylori* infections (Ma et al. 1998, You et al. 2000, Limburg et al. 2001, Wang et al. 2001, Brown et al. 2002, Wang et al. 2002, Cai et al. 2003, Wong et al. 2004), cervical cancer with Human Papillomavirus (HPV) (Zhang et al. 1989, Peng et al. 1991, IARC 1995) and liver cancer due to Hepatitis B Virus (HBV), Hepatitis C Virus (HCV) and dietary aflatoxin exposure (Chen 1993, Qian et al. 1994, Ma et al. 1998, Campbell and Chen 1999, Wang et al. 2001, Brown et al. 2002, Evans et al. 2002). The increased availability and consumption of a more diversified diet (eating considerably more fruit, vegetables and all sorts of animal products) (FAO-UN 2005), the substantially better food storage and transport, which increases the availability of fresh food and reduces the need for pickled, salted or nitrified food preservation, may have contributed to the statistically significant decline in cancers due to nutritional deficiency, such as cancers in the oesophagus, stomach and nasopharynx (Gao et al. 1994, Guo et al. 1994, Blot et al. 1995, Lu and Shi 1996, Ji et al. 1998, Kumagai et al. 1998, Yokokawa et al. 1999, Wu et al. 2000, Yang 2000, Setiawan et al. 2001, Munoz et al. 2002, Mahady et al. 2003). On the other hand, these lifestyle changes, including an increasing intake of a nutritionally richer diet (and alcohol consumption), reduction in physical activity, and overweight (especially among the young), especially in urban areas, are probably responsible for the increasing incidence and mortality of cancers of the colon-rectum, breast and prostate (Chow et al. 1993, Campbell and Chen 1999, Zhuo and Watanabe

1999, Matthews et al. 2001, Popkin 2001, Shu et al. 2001, Hao et al. 2002, Luo and Hu 2002, You et al. 2002, Zhang et al. 2002). The increasing trends for female breast cancer may also be caused by earlier menarche, which is indirectly related to greater consumption of dietary fat, high blood cholesterol (Campbell and Chen 1999), delay in the age at first childbirth, and the decline in the number of births (Hesketh and Zhu 1997, UN-SD 2002). Although cervical cancer has decreased dramatically in recent decades, it is notable that the mortality declines mainly occurred in older generations of women (Jin et al. 1999) and ceased or even reversed among younger women, trends that are more obvious in urban rather than rural areas. From a study in Shangdong Province, Li et al. (2000) proposed that the decreasing mortality from cervix uteri cancer was largely a cohort phenomenon and the increased rates among younger women reflect rapid changes in sexual mores, with increasing high-risk sexual behaviour, and greater prevalence of infection with HPV and other sexually transmitted agents (Liu et al. 1998, NSTDSCG 1998, Wang and Gao 2000).

Cigarette smoking is a major risk factor for several cancers, including lung, bladder, oesophagus, stomach, liver and cervix (IARC 2004) and the hazards of tobacco are similar for both sexes. At present, China consumes more cigarettes than any other country in the world. While comprising 20% of the world's population, China hosts 30% of the world's smokers. According to the report from the Chinese Academy of Medical Sciences, of the 1.3 billion people who live in China, 66.9% of adult men and 3.2% of adult women were smokers; in total 0.35 billion smokers over age 15 and an increasing trends appeared among young women (Yang 2004). During the last two decades, the annual consumption of cigarettes in China increased from 500 billion in 1980 to 1,800 billion in 1996; two thirds of men now become smokers before age 25. The average daily cigarette consumption by males was 1 in 1952, 4 in 1972 and 10 in 1992 (Liu et al. 1998, CTSU 1998). Of the many concerns about tobacco use in China, one of the most striking is reports that so few people understand the health consequences of smoking. In a 1996 Chinese national survey, 61% of those questioned responded that cigarette consumption posed no harm to their health (Yang 2004). Tobacco-related disease and death is an obvious consequence of China's high smoking prevalence and lack of understanding about the health effects of smoking. This epidemic of cigarette smoking in China is certainly of major importance in determining past and future trends in cancer incidence and mortality (Gao et al. 1994, Ji et al. 1996, Ji et al. 1997, Chen et al. 1997, Niu et al. 1998, Liu et al. 1998; CTSU 1998). The study conducted in 24 cities and 74 rural counties by Dr Liu and his colleagues (1998) suggested that tobacco-related deaths, estimated to be 1 million in the year 2000, would increase to 2 million/year by 2035, and 3 million/year by 2050. Smoking currently causes 12% of adult male deaths (0.6 million deaths in 1990 and 0.8 million in 2,000).

Among the tobacco-related deaths, 45% were from chronic lung disease, 15% from lung cancer and 5–8% from each of cancers in oesophagus, stomach and liver, stroke, heart disease and tuberculosis. Furthermore, current incidence rates in China reflect smokers who began their habit in the 1970s. Therefore, the increasing trends of incidence and mortality for most tobacco-related cancers will continue during next decades.

Chinese women have a high incidence of lung cancer despite a low smoking prevalence (Gao et al. 1987). According to the 1996 national prevalence survey for smoking in China: smoking remains more common among men (63% current smokers) than women (4%) (Yang et al. 1999). Most female lung cancer patients, particularly with adenocarcinoma, were life-long non-smokers. Of nonsmoking Chinese women, more than 50% were exposed to environmental tobacco smoke (ETS) at least 15 minutes per day on more than 1 day per week (Yang et al. 1999). The association between ETS and lung cancer risk in non-smoking women, and with a significant trend as increasing exposure has been confirmed in other studies conducted in urban and rural areas (Zheng et al. 1997, Zhong et al. 1999b and Wang et al. 2000). As cigarette smoking only accounted for about one-fourth of all newly diagnosed cases of lung cancer, investigations have addressed the influence of various risk factors on the incidence and mortality of lung cancer in nonsmoking Chinese women. Many studies have successfully demonstrated indoor air pollution from Chinese-style cooking, especially cooking unrefined rapeseed oil at high temperatures in woks (Gao 1996, Wang et al. 1996, Zhong et al. 1999a, Zhou et al. 2000, Matayer et al. 2002). Indoor air pollution from the solid fuel combustion, biomass fuel has been also considered as a risk for the lung cancer occurrence in many developing countries (Smith and Mahta 2003, Zheng and Smith 2003). In rural areas, an increased risk for lung cancer may also relate to the percentage of time that coal was used indoors over the past 30 years (Kleinerman et al. 2002). Several studies conducted in the cities of north-eastern China, where heavy industry is concentrated, showed increased risks of lung cancer associated with occupational exposures, such as the manufacture of transportation equipment, metal smelting, treatment and manual spinning of asbestos operation (Wu-Williams et al. 1990, Wu-Williams et al. 1993, Xu et al. 1996, Sun et al. 2003). All these findings suggested that factors other than smoking are responsible for the high risk of lung cancer among Chinese women, and it is changes in such exposures which must be evoked to explain the rising mortality and incidence trends.

Effective vaccines are available against hepatitis B infection and the protective efficacy against the development of disease or the carrier state is often 95% to 99% in cohorts of immunized infants (WHO 2002b). Two trials of vaccination against hepatitis B were set up – one in Africa (GHIS 1987) and one in Qidong County, Jiangsu Province (Sun et al. 1991) – to quantify the effectiveness in preventing liver cancer. So far they

have shown that vaccination effectively prevents the chronic carrier state. A reduction of the incidence of hepatocellular carcinoma in children was observed in Taiwan 8 years after the introduction of mass vaccination of newborns against HBV (Chang et al. 1997). Vaccines against other viral infections are currently under development. Phase II studies of vaccines against the oncogenic HPV types 16 and 18 have demonstrated a high degree of efficacy in preventing infection with these viruses, and the development of cervical neoplasia (Koutsky et al. 2002, Harper et al. 2004, Villa et al. 2005). Large phase III efficacy trials are ongoing and may confirm and extend the preliminary results. Since HPV 16 and 18 are the dominant types involved in cervical cancer etiology in all parts of the world, including China, a vaccine against these two types should, in theory, be able to prevent some 70% of cases of cancer (Muñoz et al. 2004).

Those programmes of early detection, screening and the associated effective treatment which offer the most realistic approach to reducing mortality, however, have been introduced and implemented by local health departments and are mostly confined to special occupational groups in urban areas, due to the high costs involved (Wu 1997). The efficacy and cost-effectiveness of different approaches to early diagnosis of breast cancer is of major importance. Screening programmes based on Pap smear cytology may have contributed to the declining trends of cervical cancer, along with more effective and available treatments that improved survival. Screening based on breast self-examination, according to a randomized trial in textile mills in Shanghai was shown to be ineffective in reducing mortality (Thomas et al. 2002). A major multi-centre breast cancer screening trial is now being carried out for 1 million women in China from year 2005, by the Chinese Anti-cancer Society collaborating with the American Cancer Society. Screening methods based on balloon cytology (Liu et al. 1994, Yokokawa et al. 1999) and endoscopic examination (Dawsey et al. 1998) for oesophageal cancer remained an essentially experimental approach. Other screening programmes for colorectal cancer, prostate cancer, lung cancer, and stomach cancer are even less feasible at present because of doubtful cost-effectiveness and lack of appropriate resources.

During the last decade, population growth and ageing were responsible for the biggest component of changes in the number of both cases and deaths for most cancers. In the last 50 years, the Chinese population has increased dramatically, from 550 million in 1950 to 1.24 billion in 2000, and is projected rise to 1.43 billion in 2020 (UN-PD 2003). The proportion of elderly (aged 65 or over) in the population has increased steadily: from 7.5% in 1950 to 8.6% in 1990 and to 10.1% in 2000, and is projected to be 19.5% in 2025 and 29.9% in 2050. This dramatic increase in the size and age of the population is responsible for the continuing increase in the number of cancer cases and deaths, even for cancers with decreasing trends in incidence rates, such as stomach

cancer. And with the huge population base in China, even those cancers with declining incidence and mortality, like cancers of the stomach and oesophagus, will remain a significant burden; stomach cancer will be the third most common malignancy and oesophageal cancer the fourth in 2005.

9.4 Challenges for cancer registration in China

For developing prevention strategies and monitoring the impact of control programmes, establishment of a surveillance system for cancer and its risk factors is essential and fundamental. Accurate health information would help to making better policy and to monitor change in response to cancer prevention and control, and there the cancer registry plays a very important role (Ghaffar et al. 2004). Through the result of the first national survey on cancer registry practice presented in this thesis, the enormous problems to be faced to population-based cancer registries in China are shown – related to technical, economic and other external factors. The survey suggested that lack of qualified personnel and insufficient funding support are major problems in carrying out registration work. Problems in registration methodology are also apparent from the survey, with problems in using personal identifiers, unifying the definition for the date of incidence and most valid basis of diagnosis, recording morphology and behaviour information, ICD-O coding system, and using and collecting high qualified follow-up information. The other remarkable challenge confronting cancer registries in China is the lack of stability of the population due to the huge increasing ‘floating population’ which makes registration more difficult when allocating place of residence, carrying out follow-up, and obtaining accurate denominator data, especially in urban areas.

In order to improve the quality of cancer registry data, and to compare the data between different registries in China and with other countries, several suggestions are proposed according to the results of the survey. They include recruiting more personnel, providing more systematic training about the principles and methods of cancer registration, standardizing procedures for quality control, enhancing government awareness to obtain more policy and financial support, and getting more technical support from domestic and international organisations. A directive from central or local government declaring cancer to be a notifiable disease would be a valuable contribution to cancer registration in China, but it would involve several political, financial and other administrative considerations. The network of cancer registration in China should improve the monitoring of two important components of the Chinese cancer control programme, the influence of preventive measures on the incidence trend and the cure rate from cancer measured by population-based survival, which reflects the average

prognosis of cancer in China. Some of the most successful and productive registries in the world collect only a limited amount of data for each patient (Jensen et al. 1990, Jensen et al. 1991). Therefore, the emphasis of a newly established cancer registry in China should be on the quality of the data collected, rather than on the quantity. It was suggested that most registries in China should collect only the basic information common to all registries which including subject identification (e.g. age and sex), ethnicity, incidence date, site and histology of the tumour, and the most valid basis of diagnosis. Other very useful items, such as the extent of disease (stage) and disease outcome for survival could also be considered. A unified norm or manual operation for cancer registration was suggested to be provided by the national cancer registries centre, more frequent training in coding, data management and quality control procedures, and registration software (using CanReg package developed by IARC/WHO) were also suggested.

9.5 Prospects for cancer control in China

Enormous progress in the fight against cancer has been made during the last century, much of which was directly contributed by epidemiologists, who provided the evidence for contemporary prevention strategies, through different researches or methods. A comprehensive national cancer programme (NCCP), as described by WHO (2002) evaluates the various ways to control disease and implements those that are the most cost-effective and beneficial for the largest part of the population. At the core of this cancer control strategy, the essential package includes cost-effective interventions for the following components: tobacco control, infection control, healthy eating, a curable cancer programme and palliative care. But the NCCP must also provide for information on planning, implementation, management and evaluation to help policy-makers and programme managers make the most efficient use of available resources.

In China, awareness of the importance of cancer prevention and control, especially the development of a basic information system such as cancer registration, has been increasingly addressed by government. It was emphasised by the government that the cancer registration network should be standardized and expanded in order to enhance the availability of accurate data for estimating cancer burden in China and distributing health resources efficiently. In 1990, NOCPC set up the Chinese Cancer Registry Coordination Committee, which has developed national norms and a manual for cancer registration in China, and aims to establish a national cancer registry network. The National Center for Cancer Registries was then set up in 2002, to work together with NOCPC. At the end of 2003, the Ministry of Health in China approved and circulated

an official document about the 'Framework of Cancer Prevention and Control in China (2004–2010)' (MOH-China 2003). This framework aims to enhance the prevention and control of cancer in China so as to minimize the threat to the Chinese people from cancers, improve public health conditions in general, and to promote the people's welfare in accordance with their socio-economic development. A detailed implementation project was further developed, covering four major tasks: anti-cancer education, comprehensive and pertinent cancer information system establishment, cancer risk factors surveillance, and cancer diagnosis and treatment improvement (including a cancer screening programme).

A comprehensive vital registration system plays an important role in monitoring cancer prevention and control programmes. To date, there has been no national vital registration system in China. The CHIS vital registration system covers the largest population, and is the source of mortality data for the WHO statistical information system (WHOSIS), but more effort is needed to improve completeness, coverage, validity and accuracy to get a representative profile of national mortality. The DSP system, as shown in this study, and in a recent assessment completed by China CDC (Setel et al. 2005, Yang et al. 2005), has weaknesses when used to estimate national mortality. In consequence, in 2004 a systematic adjustment of DSP was conducted by China CDC, to improve the representativeness. A revised DSP system will include 150 sample points, expanded to cover 6% of whole population, with about 1/3 of the sites being newly selected. The DSP system was initially designed as a mortality registry for surveillance of the major infectious diseases, but this function has become redundant since a national infectious disease notification system was established in China in 2004. Currently, the DSP system focuses on non-communicable disease surveillance. (Yang et al. 2005, Wang et al. 2005) Based on the experiences from the two former national mortality surveys in the 1970s and the 1990s, the MOH in China has recently launched the third national mortality survey in 2005. Coordinated by the MOH, NOCPC and CDC are collaborating to conduct the whole survey during the next two years. It is a 10% sample survey, based mainly on the sample units used in the second mortality survey. However, both DSP and CHIS data sample units will be considered for inclusion in the survey. Information on deaths occurring in the sample populations during the years 2003–2005 will be collected retrospectively. The aim is to establish the current mortality profile in the country, and the changes that have occurred during last 30 years will be evaluated by comparisons with the earlier two national mortality surveys. The third survey will also, it is hoped, provide the foundations for a comprehensive and representative vital registration system based upon the different ongoing death reporting systems in the country.

One proposal for enhancing Chinese cancer registration work was to set up a stratified cancer registry network according to local conditions, aiming to cover 10% of the national population, and to develop a Chinese cancer incidence database. Different cancer registration training classes would be held in the country, focusing on cancer registration principles, methods, coding systems, software, data analysis and application. For establishing a systematic surveillance of the prevalence for major risk factors for cancer and other NCD's, WHO SuRF Reports (WHO 2005d) has already summarized the major risk factor profiles for NCD in China, based on national sampling or local level surveys in the country, such as the DSP system. In recent decades, some other large scale studies were carried out focusing on cancer risk factor surveillance, such as the big retrospective study about tobacco hazards organized by CAMS, CAPMS and CTSU in the UK during 1989–1991 (Liu et al. 1998, CTSU 1998); the big sample surveys done in the context of a large research project organised by CAPMS in collaboration with Cornell and Oxford Universities in 65 Chinese counties about 'diet, lifestyle and mortality in China' from the late 1980s (Chen et al. 1990). Future work will likely be focused on ongoing risk factor surveillance in selected populations, based upon the methodology of WHO STEPwise approach for Surveillance of noncommunicable diseases (STEPS). This involves three sequential steps in surveillance for cancer risk: questionnaire data, objective physical measurements and then biochemical measurements. (WHO 2005c) The initial sample places could either focus on some big cities (such as Beijing, Tianjin, Shanghai) or continue with the sites used in previous studies. A cancer risk factor information database could be set up. Tobacco control remains the most important strategy for preventing cancers. The Framework Convention on Tobacco Control that the member states of WHO agreed to submit to the World Health Assembly in May 2003 represented a powerful tool to ensure that such strategies are implemented. The Ministry of Health (MOH) in China will follow this framework and the anti-smoking campaign was advocated in more areas. WHO is also engaged in preparing a Global Strategy on Diet, Physical Activity and Health, due to the growing global burden of chronic diseases such as cancer, cardiovascular disease, diabetes and obesity. More consumption of locally produced vegetables, fruit and agricultural products will be encouraged, while high fat and high cholesterol consumption should be avoided.

Vaccination against HBV was introduced nationwide into routine infant immunization from April 2002 (MOH-China 2005), but will not impact on incidence and mortality rates for many years. As great progress shown in Phase II trials (Koutsky et al. 2002, Harper et al. 2004, Villa et al. 2005), vaccination against cervical cancer will, undoubtedly, be the chosen prevent method for cervix cancer in the long term. Nevertheless, the impact on the rates of cervical cancer would appear even later.

So far, no mass screening programme exists in China, although there has been research on the efficacy or cost-effectiveness of screening programmes for certain cancers. Recently, the MOH in China organized some experts to develop guidelines for screening programmes for certain common cancers in China. However, to date the major focus remains upon opportunistic screening, early detection of cancer, and research into methodology. It is obvious that, with the demand request for early detection or screening for cancer, it is not feasible to expect nation-wide policy implementation. Nevertheless, national recommendations are needed, analogous to those prepared by the US Preventive Services Task Force (USPSTF 1996), while continuing research into the most appropriate methods for early detection in the Chinese context. A Task Force could be a forum for reviewing the results of screening projects specific to the Chinese situation, and a screening committee would be a useful approach to review and supervise ongoing projects on screening in China. This committee intends to make national level recommendations for screening policy and research and be a definite advice centre in China – providing suggestions to the MOH in China. As results from Swedish randomized controlled trials on mammography screening (Nystrom et al. 2002), a significant 21% reduction in breast cancer mortality (RR=0.79, 95% CI 0.70–0.89) was found between screening and non-screening women aged 40–74 and after long-term follow-up (20 years) in two counties, the screening decreased breast cancer mortality by 32% in screened women aged 40–69. Some anti-cancer strategies have been proposed, such as Taihang Anti-Cancer Campaign for esophageal cancer control in the high-risk area along Taihang Mountains, and 1 million women breast cancer screening programme in China (Qiao et al. 2001, Dong et al. 2002). Nevertheless, it is necessary to have estimation on cost-effectiveness of such interventions before proposing widespread implementation.

Although the NOCPC has played an important role in prevention strategies, the increasing cancer burden facing China and the multidisciplinary cooperation in a complicated and integrated cancer prevention and control programme, it is obviously too small for the big tasks it was given. Currently, a National Cancer Center (NCC) is proposed to be established in Beijing, which would cover all of the work currently done by NOCPC, plus some extra functions, such as cancer screening, surveillance and education about cancer prevention and control.

10. Conclusions

Implementing a programme of cancer prevention and control requires resources and processes, all of which have to be well managed, and based on an accurate knowledge of the cancer burden in the country. Currently there is no well-established disease surveillance mechanism providing precise or representative estimation for the whole Chinese population. However, based on the most recently available representative national mortality survey (1990–1992), following the time trends from the routine mortality reporting system (CHIS data) and the relationship between incidence and mortality as observed in cancer registration data (7 registries data from CI5 the 8th version), this thesis reports the results of an analysis of current cancer incidence and mortality pattern on a national scale.

The size and direction of change is clear – the cancer burden is rising in China. There are large increases in the absolute number of cancer cases and deaths, not only because of the overall increases in risk of the disease, but even more importantly, because of huge population growth and aging in the country. The leading cancers in China in 2000 were cancers in lung, stomach, liver, oesophagus and large bowel in men, and stomach, breast, lung, liver, oesophagus, large bowel and uterine cervix in women. There are wide variations between urban and rural areas, so the most common cancers differ between different populations. Steady increases in rates of cancers in lung and female breast and decreases in cancers of the stomach and oesophagus are seen in the whole country, and these trends are expected to continue in the future. A large part of the dramatic increase in cancer incidence and mortality is due to a rise in lung cancer, attributable to the increasing prevalence of cigarette smoking for both sexes and projected to become increasingly common over the next decades, even if anti-smoking campaigns succeed in promoting smoking cessation, as experience has shown in the western world (Storm et al. 2003, NCI-SRAB 2005). Controlling the tobacco epidemic is the greatest long-term challenge for public health in China at the beginning of the 21st century (Liu et al. 1998, Niu et al. 1998).

To reduce cancer incidence and mortality, and to improve the quality of life of people with cancer, a comprehensive prevention and control programme (i.e. NCCP), which includes strengthened surveillance and management programmes, needs to be developed and effectively carried out in China, according to the local context in the country, with reference to experiences and lessons from other countries. Establishing a national cancer risk factor surveillance system would provide a sound platform for monitoring and controlling the major risk factors related to cancer. Risk factor

surveillance would help to focus the appropriate emphasis to be placed on different primary prevention strategies for cancer, such as anti-smoking campaigns, proposing healthy diet, controlling infections, such as HBV and HPV, increasing physical activity, eliminating occupational exposures and so on. As for the secondary prevention strategy, it is suggested that national recommendations for cancer screening, early diagnosis and treatment programmes be developed based on the various studies on the feasibility and cost-effectiveness, related to the local context in the country, and also the successful experiences from other countries.

Estimation and projection of the cancer burden is clearly an essential step in planning and allocation of resources, but the methods used in our study must be considered as only a surrogate for measuring cancer pattern by means of the systematic mortality surveillance and registration of cancer cases arising in the population. A surveillance and monitoring system is the basis for any rational programme of cancer control. This should include improvement and extension of the existing system of death registration, to improve its quality, and representativeness of the national population. A national representative cancer registration network will provide accurate information and increase the validity of the estimates on the burden and pattern of cancer in the community as evidence-based setting priorities for the NCCP, including the planning of cancer-related health care services, as well as promoting a solid basis for research on cancer etiology, prevention and monitoring the effects of early detection/screening and treatment. Registration of cancer incidence and mortality in China should and would be standardized and expanded, as the progress shown in the country. Nevertheless, for the moment, with various barriers due to technical, financial or political reasons in the region, the national representative network remains a long-term aspiration. In the meantime, we believe that the estimates based on the results of the current mortality surveillance programme and the few existing registries with high-quality data provide a fair overview of the problem, and indicate the priorities for cancer control at the national level in China.

11. Acknowledgements

This study was carried out in the International Agency for Research on Cancer (IARC/WHO) in Lyon, France, the Tampere School of Public Health in Finland and the National Office for Cancer Prevention and Control in China, and has been supported by all above three organisations.

First of all, I wish to express my deepest gratitude to my supervisor, Dr. Donald Maxwell Parkin who was the leader of the Cancer Epidemiology & Biostatistic Cluster in IARC/WHO when I was working there and supervised me throughout this study during last 4 years. I am very honoured to have been supervised by Dr. Parkin and appreciated his broad and profound professional expertise and knowledge (not only in cancer epidemiology). His critical insight and enthusiastic attitude towards science, the optimistic, warmhearted and positive attitude towards life indeed taught and influenced me a lot. And, his perfectly used native language really improved my speaking and writing English remarkably.

I want to express my warmest thanks to my another supervisor, Professor Anssi Auvinen, especially for his excellent teaching, marvellous knowledge of epidemiology and his great help in my thesis writing.

My appreciation is also due to the two pre-examiners of this dissertation: Professor Timo Hakulinen, Director of the Finnish Cancer Registry, especially for his tremendous knowledge of statistics, critical and useful comments and warmhearted help, and to Professor Alan Lopez in the University of Queensland, Australia. I also wish to thank two professors on my Evaluation Committee, Matti Hakama and Risto Sankila of the Tampere School of Public Health, for their professional expertise and friendly helps to my work.

My warmest thanks also to the people in China, especially to Professors Ping Zhao, Liandi Li, Yude Chen, Mr. Siwei Zhang and all my other colleagues in the National Office for Cancer Prevention and Control and the Center of Health Statistics and Information in the Ministry of Health in China. Without their great support and hard work in China, this work could not have been accomplished. My thanks also to Dr. Robert Burton, a WHO Consultant in China, for his kindly help and support.

My warmest thanks also to all my former colleagues in the IARC/WHO in Lyon, France, especially Sharon Whelen, Jacques Ferlay, Freddie Bray, Chantal Déchaux, Krittika Pitaksaringkarn, for their great kindness helps to this study and their good

friendships. Also my great thanks to all my friends and colleagues there, especially Esther, Anja, Min, Mia and Maisa who gave me two wonderful years in my life there. It has been a great pleasure for me to have been working in that international agency and feeling the warm atmosphere of the whole world's family.

I wish to express my great thanks to all teachers, staff and all my former classmates in the IPPE group at the Tampere School of Public Health, especially to my Finnish sister, Catarina, for her great and warmhearted supports and helps and my good friends Grace, Rex and Yanru.

My deepest gratitude is to my parents Guo-song and Zhong-zhen, my three elder sisters Ping, Li, Hui and their families, for their endless love and forever supporting and encouraging me all my life.

Finally, my warmest and deepest appreciation is to my dear husband, Guangli, whose great love, understanding, patience and endless supporting is indeed the best treasure I have in this world.

A handwritten signature in black ink, reading "Ling Yang". The signature is written in a cursive, flowing style with a large initial 'L' and 'Y'.

Ling Yang,
Oct. 2005, from Tampere

12. References

- Alderson M (1981): International Mortality Statistics. Facts On File Inc. New York, USA
- American Cancer Society (ACS) (2005, copied on 15th April):
http://www.cancer.org/docroot/STT/STT_0.asp
- Bhurgri Y, Bhurgri A, Hassan SH, Zaidi SHM, Rahim A, Sankaranarayanan R and Parkin DM (2000): Cancer incidence in Karachi, Pakistan: First results from Karachi Cancer Registry. *Int J Cancer* 85:325–329
- Black RJ, Bray F, Ferlay J and Parkin DM (1997): Cancer incidence and mortality in the European Union.: cancer registry data and estimates of national incidence for 1990. *Eur J Cancer* 33:1075–1107
- Blot WJ, Li JY, Taylor PR, Guo W, Dawsey SM and Li B (1995): The Linxian trials: mortality rates by vitamin-mineral intervention group. *Am J Clin Nutr* 62 (6 Suppl):1424s–1426s
- Bray F, Sankila R, Ferlay J and Parkin DM (2002): Estimates of cancer incidence and mortality in Europe in 1995. *Eur J Cancer* 38:99–166
- Breslow NE (1984): Extra-Poisson variation in log-linear models. *Applied Statistics* 39:38–44
- Brown CC and Kessler LG. (1988): Projections of lung cancer mortality in the United States: 1985–2025. *J. Natl. Cancer Inst.* 80:43–51
- Brown LM, Thomas TL, Ma JL, Chang YS, You WC, Liu WD, Zhang L, Pee D and Gail MH (2002): Helicobacter pylori infection in rural China: demographic, lifestyle and environmental factors. *Int J Epidemiol* 31:638–646
- Cai L, Zheng ZL and Zhang ZF (2003): Risk factors for the gastric cardia cancer: a case-control study in Fujian Province. *World J Gastroenterol* 9:214–218
- Campbell TC and Chen JS (1999): Energy Balance: Interpretation of data from rural China. *Toxicological Science* 52(Suppl): 87–94
- Capocaccia R, Verdecchia A, Micheli A, Sant M, Gatta G and Berrino F (1990): Breast cancer incidence and prevalence estimated from survival and mortality. *Cancer Causes Control* 1, 23–29
- Center of Health Information and Statistical in the Ministry of Health (CHIS-MOH) (2002, released in May 2002): The abstract for China Health Care in 2001:
<http://www.moh.gov.cn/statistics/digest01/index.htm>
- Centers for Disease Control and Prevention in United States of America (CDC-USA) (2005, copied on 15th April): <http://www.cdc.gov/nchs/about/major/dvs/mortdata.htm>
- Centers for Disease Control and Prevention in US (CDC-USA) (2005, last reviewed in March): <http://www.cdc.gov/nchs/about/major/dvs/mortdata.htm>
- Chang MH, Chen CJ, Lai MS, Hsu HM, Wu TC, Kong MS, Liang DC, Shau WY and Chen DS (1997): Universal hepatitis B vaccination in Taiwan and the incidence of hepatocellular carcinoma in children. Taiwan Childhood Hepatoma Study Group. *N Engl J Med* 336:1855–1859.
- Chen JG (1993): Epidemiologic study of liver cancer in the past 20 years. *Zhonghua Liu Xing Bing Xue Za Zhi* 14:241–243.
- Chen JS, Campbell TC, Li JY and Peto R (eds) (1990): Diet, lifestyle and mortality in China. Oxford University Press, Oxford, Cornell University Press, Ithaca, and People's Publishing House, Beijing
- Chen KX, Wang PP, Zhang SW, Li LD, Lu FZ and Hao XS (2003): Regional variations in mortality rates of pancreatic cancer in China: results from 1990–1992 national mortality survey. *World J Gastroenterol.* 9: 2557–2560

- Chen Z, Zheng T and Chen J (2000): Evaluation of ten-year results of cancer prevention and treatment in Changle City with high incidence of gastric cancer. *Zhonghua Zhong Liu Za Zhi* 22(4):311–313
- Chen ZM, Xu Z, Collins R, Li WX and Peto R (1997): Early health effects of the emerging tobacco epidemic in China. A 16-year prospective study. *JAMA* 278: 1500–1504
- Chow WH, Dosemeci M, Zheng W, Vetter R, McLaughlin JK, Gao YT and Blot WJ (1993): Physical activity and occupational risk of colon cancer in Shanghai, China. *Int J Epidemiol* 22: 23–29
- Clayton D and Schifflers E (1987a): Models for temporal variation in cancer rates. I. Age-period and age-cohort models. *Stat Med* 6: 449-467
- Clayton D and Schifflers E (1987b): Models for temporal variation in cancer rates. II: Age-period-cohort models. *Stat Med* 6: 469-481
- Clinical Trial Service Unit (CTSU) (1998): *Emerging Tobacco Hazards in China*. <http://rum.ctsu.ox.ac.uk/tobacco>. Oxford University Press, Oxford, UK
- Coleman MP, Esteve J, Damiecki P, Arslan A and Renard H (1993): Trends in cancer incidence and mortality. Lyon IARC Sci Publ No. 121: 27–41
- Dawsey SM, Fleischer DE, Wang GQ, Zhou B, Kidwell JA, Lu N, Lewin KJ, Roth MJ, Tio TL and Taylor PR (1998): Mucosal iodine staining improves endoscopic visualization of squamous dysplasia and squamous cell carcinoma of the esophagus in Linxian, China. *Cancer* 83:220–231
- Deerasamee S, Martin N, Sontipong S, Sriamporn S, Sriplung H, Srivatanakul P, Vatanasapt V, Parkin DM and Ferlay J (eds) (1999): *Cancer in Thailand Vol. II 1992–1994*. IARC Technical Report No. 34, IARCPress, Lyon, France
- Dept of Control Disease of MOH, Chinese Academy of Preventive Medicine (DCD-MOH and CAPM) (eds) (1999): *A series of Reports on Chinese Disease Surveillance (8) 1997 Annual Report on Chinese Disease Surveillance*. Beijing Medical University Press, Beijing, China
- Devesa SS, Donaldson J and Fears T (1995): Graphical presentation of trends in rates. *Am J Epidemiol* 141: 300–304
- Dickman PW, Sloggett A, Hills M and Hakulinen T (2004): Regression models for relative survival. *Stat Med* 23:51–64
- Doll R (1960): The geographical distribution of cancer. *Br J Cancer* 23: 1–8
- Doll R, Muir C and Waterhouse J (1970): *Cancer Incidence in Five Continents, Vol. II*. Geneva, UICC; Berlin, Springer-Verlag
- Doll R, Payne P and Waterhouse J (1966): *Cancer Incidence in Five Continents, Vol. I. A Technical Report*. Geneva, UICC; Berlin, Springer-Verlag
- Dong ZW, Tang PZ, Li LD and Wang GQ (2002): The strategy for esophageal cancer control in high-risk areas of China. *Jpn J Clin Oncol* 32 Suppl: S10–12
- dos Santos Silva I (1999): *Cancer Epidemiology: Principles and Methods*. IARCPress, Lyon, France
- Dyba T and Hakulinen T (2000): Comparison of different approaches to incidence prediction based on simple interpolation techniques. *Stat Med* 19: 1941–1952
- Dyba T, Hakulinen T and Paivarinta L (1997): A simple non-linear model in incidence prediction. *Stat Med* 16: 2297-2309
- Editorial Committee for the Atlas of Cancer Mortality (eds) (1979): *Atlas of cancer mortality in the People's Republic of China*. China Map Press, Beijing, China
- Engeland A, Haldorsen T, Tretli S, Hakulinen T, Horte LG, Luostarinen T, Magnus K, Schou G, Sigvaldason H, Storm HH, Tulinius H and Vaittinen P (1993): Prediction of cancer incidence in the Nordic countries up to the years 2000 and 2010. *Acta Path Microbiol Scand* 101(Suppl) 38: 1–124
- Engeland A, Haldorsen T, Tretli S, Hakulinen T, Horte LG, Luostarinen T, Schou G, Sigvaldason H, Storm HH, Tulinius H and Vaittinen P (1995): Prediction of cancer mortality in the Nordic countries up to the years 2000 and 2010. *Acta Path Microbiol Scand* 101(Suppl) 49: 1–161

- Estève J, Benhamou E and Raymond L (eds) (1994): Statistical methods in cancer research. Vol. IV. Descriptive epidemiology. International Agency for Research on Cancer, Lyon. IARC Sci Publ No. 128:170–206
- Estève J, Kricke A, Ferlay J and Parkin DM (eds) (1993): Facts and figures of cancer in the European Community. International Agency for Research on Cancer. Lyon, France
- European Network of Cancer Registries (ENCR) (1995): EUROCIM User Manual, 2nd edn. Lyon, International Agency for Research on Cancer, 1995
- European Network of Cancer Registries (ENCR) (2001): EUROCIM version 4.0. European Incidence Database V2.3. Lyon, France: IARCPress; 2001.
- European Network of Cancer Registries (ENCR) (2005, modified on 13th April): <http://www.enrc.com.fr/>
- Evans AA, Chen G, Ross EA, Shen FM, Lin WY and London WT (2002): Eight-year follow-up of the 90,000-person Haimen City cohort: I. Hepatocellular carcinoma mortality, risk factors, and gender differences. *Cancer Epidemiol Biomarkers Prev.* 11: 369–376.
- Eylenbosch WJ and Noah ND (1988): Surveillance in Health and Disease. Oxford Univ. Press. Oxford, UK
- Ferlay J, Bray F, Pisani P and Parkin DM (2001): GLOBOCAN 2000: Cancer Incidence, Mortality and Prevalence worldwide, Version 1.0. IARC Cancer Base No.5, IARCPress, Lyon, France
- Ferlay J, Bray F, Pisani P and Parkin DM (2004): GLOBOCAN 2002: Cancer Incidence, Mortality and Prevalence Worldwide, Version 2.0. IARC CancerBase No. 5, IARCPress, Lyon, France
- Ferlay J, Bray F, Sankila R and Parkin DM (1999): EUCAN: Cancer Incidence, Mortality and Prevalence in the European Union 1995, version 2.0. IARC CancerBase No. 4. Lyon, IARCPress, Lyon, France
- Fernandez E, Gonzalez JR, Borrás JM, Moreno V, Sanchez V and Peris M (2000): Recent decline in cancer mortality in Catalonia (Spain). A joinpoint regression analysis. *Eur J Cancer* 37: 2222–2228
- Food and Agriculture Organization of the United Nations (FAO-UN) (2005, copied on 15th April): FAOSTAT: <http://apps.fao.org/faostat/collections?version=ext&hasbulk=0>
- Francis B, Green M and Payne C (1993): The GLIM System. Release 4 Manual. Clarendon Press, Oxford, UK
- Gao C (1991): The mortality of patients with malignant tumors in eleven cities of Jiangsu Province, 1984–1986. *Zhonghua Liu Xing Bing Xue Za Zhi*;12:158–162
- Gao YT (1996): Risk factors for lung cancer among nonsmokers with emphasis on lifestyle factors. *Lung Cancer* 14 Suppl: S39–45
- Gao YT, Blot WJ, Zheng W, Ershow AG, Hsu CW, Levin LI, Zhang R and Fraumeni JF Jr (1987): Lung cancer among Chinese women. *Int J Cancer* 40: 604–609
- Gao YT, McLaughlin JK, Blot WJ, Ji BT, Benichou J, Dai Q and Fraumeni JF Jr (1994): Risk factors for esophageal cancer in Shanghai, China. I. Role of cigarette smoking and alcohol drinking. *Int J Cancer* 58: 192–196
- Gao YT, McLaughlin JK, Gridley G, Blot WJ, Ji BT, Dai Q and Fraumeni JF Jr (1994): Risk factors for esophageal cancer in Shanghai, China. II. Role of diet and nutrients. *Int J Cancer* 58: 197–202
- Gao YT, Tu JT, Jin F and Gao RN (1981): Cancer mortality in Shanghai during the period 1963–77. *Brit. J. Cancer* 43: 183–195
- Ghaffar A, Reddy KS and Singhi M (2004): Burden of non-communicable diseases in South Asia. *BMJ* 328: 807–810
- GHIS (1987): The Gambia Hepatitis Intervention Study. The Gambia Hepatitis Study Group. *Cancer Res.* 47, 5782–5787
- Greenlee RT, Murray T, Bolden S and Wingo PA (2000): Cancer Statistics, 2000. *CA Cancer J Clin* 50:7–33
- Guo WD, Blot WJ, Li JY, Taylor PR, Liu BQ, Wang W, Wu YP, Zheng W, Dawsey SM, Li B and Fraumeni JF Jr (1994): A nested case-control study of oesophageal and stomach cancers in the Linxian nutrition intervention trial. *Int J Epidemiol* 23:444–450

- Hakulinen T (1996) The future cancer burden as a study subject *Acta Oncol* 35: 665–670
- Hakulinen T and Dyba T (1994): Precision of incidence predictions based on Poisson distributed observations. *Stat Med* 13:1513–1523
- Hakulinen T and Pukkala E (1981): Future incidence of lung cancer: forecasts based on hypothetical changes in the smoking habits of males. *Int J Epidemiol* 10:233–240.
- Hao XS, Chen KX and Liu H (2002): Analysis of female breast cancer descriptive epidemiology in Tianjin, China. *Asian Pacific J Cancer Prev* 3: 143–148
- Harper DM, Franco EL, Wheeler C, Ferris DG, Jenkins D, Schuind A, Zahaf T, Innis B, Naud P, De Carvalho NS, Roteli-Martins CM, Teixeira J, Blatter MM, Korn AP, Quint W, Dubin G; GlaxoSmithKline and HPV Vaccine Study Group (2004): Efficacy of a bivalent L1 virus-like particle vaccine in prevention of infection with human papillomavirus types 16 and 18 in young women: a randomised controlled trial. *Lancet* 364: 1757–1765
- He YT, Yang L, Hou J, Shao Z, Zhang SW, Meng FS, Lian SY and Qiao YL (2001): Prevalence Trends of esophageal cancer in Cixian, Hebei Province and Linxian, Henan Province. *Cancer Research for Prevention and Control* 28: 485–488
- Heasman HA and Lipworth L (eds) (1966): Accuracy of Certification of Cause of Death: Studies in Medical and Population Subjects No. 20. London, United Kingdom:
- Heilig GK (1999 web vision 1.1): Can China Feed Itself ? A system for Evaluation of Policy Options. http://www.iiasa.ac.at/research/LUC/ChinaFood/index_m.htm
- Hesketh T and Zhu WX (1997): The one child family policy: the good, the bad, and the ugly. *Brit. Med. J* 314: 1685–1687
- Heuer C (1997): Modeling of time trends and interactions in vital rates using restricted regression splines. *Biometrics* 53: 161–177
- International Agency for Research on Cancer (IARC) (1995): Monographs on the evaluation of the carcinogenic risks to humans, vol 64: human papillomaviruses. IARCPress, Lyon, France
- International Agency for Research on Cancer (IARC) (2004): Tobacco Smoke and Involuntary Smoking, vol. 83, IARC, Lyon, France
- International Agency for Research on Cancer (IARC) (2005, modified version on 11 January 2005): Cancer Mondial Project. <http://www-dep.iarc.fr>
- Jensen OM, Estève J, Møller H and Renard H (1990): Cancer in the European Community and its member states. *Eur J Cancer* 26:1167–1256
- Jensen OM, Parkin DM, MacLennan R, Muir CS and Dkeet RG (eds) (1991): Cancer registration: principles and methods. IARC Scientific Publication No. 95, IARCPress, Lyon, France
- Ji BT, Chow WH, Yang G, McLaughlin JK, Gao RN, Zheng W, Shu XO, Jin F, Fraumeni JF Jr and Gao YT (1996): The influence of cigarette smoking, alcohol, and green tea consumption on the risk of carcinoma of the cardia and distal stomach in Shanghai, China. *Cancer* 77:2449–2457
- Ji BT, Chow WH, Yang G, McLaughlin JK, Zheng W, Shu XO, Jin F, Gao RN, Gao YT and Fraumeni JF Jr (1998): Dietary habits and stomach cancer in Shanghai, China. *Int J Cancer* 76:659–664
- Ji BT, Shu XO, Linet MS, Zheng W, Wacholder S, Gao YT, Ying DM and Jin F (1997): Paternal cigarette smoking and the risk of childhood cancer among offspring of nonsmoking mothers. *J Natl Cancer Inst* 89: 238–244
- Jin F, Devesa SS, Chow WH, Zheng W, Ji BT, Fraumeni JF Jr and Gao YT (1999): Cancer incidence trends in urban Shanghai, 1972–94: An update. *Int. J. Cancer* 83: 435–440
- Kim HJ, Fay MP, Feuer EJ and Midthune DN (2000): Permutation tests for joinpoint regression with applications to cancer rates. *Stat Med.* 19: 335–351
- Kleinerman RA, Wang Z, Wang L, Metayer C, Zhang S, Brenner AV, Zhang S, Xia Y, Shang B and Lubin JH (2002): Lung cancer and indoor exposure to coal and biomass in rural China. *J Occup Environ Med.* 44:338–344
- Koutsky LA, Ault KA, Wheeler CM, Brown DR, Barr E, Alvarez FB, Chiacchierini LM and Jansen KU (2002): A controlled trial of a human papillomavirus type 16 vaccine. *N Engl J Med.* 347: 1645–1651

- Kumagai Y, Pi JB, Lee S, Sun GF, Yamanushi T, Sagai M and Shimojo N (1998): Serum antioxidant vitamins and risk of lung and stomach cancers in Shenyang, China. *Cancer Lett.* 129: 145–149.
- Li HQ, Jin SQ, Xu HX and Thomas DB (2000): The decline in the mortality rates of cervical cancer and a plausible explanation in Shandong, China. *Int J Epidemiol* 29: 398–404.
- Li LD and Rao KQ (2000) The cancer incidence and mortality statistics of 11 cancer registries in China, 1988–1992. *Bull Chin Cancer* 9:435–447
- Li LD, Lu FZ, Zhang SW, Mu R, Sun XD, Huangpu XM, Sun J, Zhou YS, Ouyang NH, Rao KQ, Chen YD, Sun AM, Xue ZF and Xia Y (1996): Analysis of cancer mortality and distribution in China from year 1990 through 1992. *An Epidemiologic Study. Chin J Oncol* 18: 403–407
- Li LD, Lu FZ, Zhang SW, Mu R, Sun XD, Huangpu XM, Sun J, Zhou YS, Ouyang NH, Rao KQ, Chen YD, Sun AM, Xue ZF and Xia Y (1997a): Analysis of variation trend and short term detection of Chinese malignant tumor mortality during twenty years. *Chin J Oncol* 19: 3–9
- Li LD, Zhang SW, Lu FZ, Mu R, Sun XD, Huangpu XM, Sun J, Zhou YS, Xia Y, Dai XD, Rao KQ, Chen YD, Sun AM and Xue ZF (1997b): The character for the spectrum and component of cancer mortality in China. *Chin J Oncol* 19: 323–328
- Limburg P, Qiao Y, Mark S, Wang G, Perez-Perez G, Blaser M, Wu Y, Zou X, Dong Z, Taylor P and Dawsey S (2001): Helicobacter pylori seropositivity and subsite-specific gastric cancer risks in Linxian, China. *J Natl Cancer Inst* 93:226–233.
- Liu BQ, Peto R, Chen ZM, Boreham J, Wu YP, Li JY, Campbell TC and Chen JS (1998): Emerging tobacco hazards in China: 1. Retrospective proportional mortality study of one million deaths. *Brit Med J* 317:1411–1422.
- Liu H, Xie J, Yu W, Song W, Gao Z, Ma Z and Detels R (1998): A study of sexual behavior among rural residents of China. *J Acquir Immune Defic Syndr Hum Retrovir* 19:80–88
- Liu SF, Shen Q, Dawsey SM, Wang GQ, Nieberg RK, Wang ZY, Weiner M, Zhou B, Cao J, Yu Y, Guo WD, Li JY, Blot WJ, Li B and Taylor PR (1994): Esophageal balloon cytology and subsequent risk of esophageal and gastric-cardia cancer in a high-risk Chinese population. *Int J Cancer* 57:775–780
- Lopez AD and Murray CJ (1998). The global burden of disease, 1990–2020 *Nature Medicine* 4: 1241–1243
- Lopez AD (2005): The evolution of the Global Burden of Disease framework for disease, injury and risk factor quantification: developing the evidence base for national, regional and global public health action. *Global Health*.1:5
- Lu JB, Yang WX, Liu JM, Li YS and Qin YM (1985): Trends in morbidity and mortality for oesophageal cancer in Linxian County, 1959–1983. *Int. J Cancer* 36: 643–645
- Lu JB, Yang WX, Zu SK, Chang QL, Sun XB, Lu WQ, Quan PL and Qin YM (1988): Cancer mortality and mortality trends in Henan, China, 1974–1985. *Cancer Detect Prev* 13: 167–173
- Lu JC and Shi LY (1996): A case-control study of childhood leukemia. *Zhonghua Liu Xing Bing Xue Za Zhi.* 17: 229–232
- Luo J and Hu FB (2002): Time trends of obesity in pre-school children in China from 1989 to 1997. *Int J Obes Relat Metab Disord.* 26: 553–558
- Ma JL, You WC, Gail MH, Zhang L, Blot WJ, Chang YS, Jiang J, Liu WD, Hu YR, Brown LM, Xu GW and Fraumeni JF Jr (1998): Helicobacter pylori infection and mode of transmission in a population at high risk of stomach cancer. *Int J Epidemiol* 27:570–573
- Mahady GB, Pendland SL, Yun GS, Lu ZZ and Stoia A (2003): Ginger (*Zingiber officinale* Roscoe) and the gingerols inhibit the growth of Cag A+ strains of Helicobacter pylori. *Anticancer Res* 23:3699–3702
- Mathers CD, Shibuya K, Boschi-Pinto C, Lopez AD and Murray CJ (2002): Global and regional estimates of cancer mortality and incidence by site: I. Application of regional cancer survival model to estimate cancer mortality distribution by site. *BMC Cancer* 2:36

- Mathers CD, Fat DM, Inoue M, Rao C and Lopez AD (2005): Counting the dead and what they died from: an assessment of the global status of cause of death data. *Bull World Health Organ* 83:171–177.
- Matthews CE, Shu XO, Jin F, Dai Q, Hebert JR, Ruan ZX, Gao YT and Zheng W (2001): Lifetime physical activity and breast cancer risk in the Shanghai Breast Cancer Study. *Br J Cancer* 84:994–1001
- Ménégoz F, Black RJ, Arveux P, Magne V, Ferlay J, Buemi A, Carli PM, Chapelain G, Faivre J, Gignoux M, Grosclaude P, Mace-Lesec'h J, Raverdy N and Schaffer P (1997): Cancer Incidence and Mortality in France in 1975–1995. *Eur J Cancer Prev* 6: 442–466
- Metayer C, Wang Z, Kleinerman RA, Wang L, Brenner AV, Cui H, Cao J and Lubin JH (2002): Cooking oil fumes and risk of lung cancer in women in rural Gansu, China. *Lung Cancer* 35:111–117
- Miñarro R, Black RJ, Martínez C, Navarro C, Garau I, Izarzugaza I, Ardanaz E, Vergara A, Galcerán J, Alonso R, Mateos A and Rodríguez M (eds) (2000): *Cancer Incidence and Mortality in Spain – Patterns and Trends*. IARC Technical Report No. 36, IARCPress, Lyon, France
- Ministry of Health in P.R. China (MOH-China) (eds) (2003): *The Framework of Cancer Prevention and Control in China (2004–2010)*. Beijing, China
- Ministry of Health in P.R. China (MOH-China) (2005, copied on 15th April): <http://www.moh.gov.cn/>
- Møller B, Fekjær H, Hakulinen T, Tryggvadóttir L, Storm HH, Talbäck M and Haldorsen T (2002): Prediction of cancer incidence in the Nordic countries up to the year 2020. *Eur J Can Prev* 11(Suppl 1):S1–96
- Muir C, Waterhouse J, Mack T, Powell J and Whelan S (eds) (1987): *Cancer incidence in five continents*. Vol V. IARC Scientific Publications No. 88, IARCPress, Lyon, France
- Muñoz N, Bosch FX, Castellsague X, Diaz M, de Sanjose S, Hammouda D, Shah KV and Meijer CJ (2004): Against which human papillomavirus types shall we vaccinate and screen? The international perspective. *Int J Cancer* 111:278–285.
- Muñoz N, Franceschi S, Bosetti C, Moreno V, Herrero R, Smith JS, Shah KV, Meijer CJ, Bosch FX and International Agency for Research on Cancer Multicentric Cervical Cancer Study Group (2002): Role of parity and human papillomavirus in cervical cancer: the IARC multicentric case-control study. *Lancet* 359:1093–1101
- Murray CJ and Lopez AD (eds) (1996): *GLOBAL BURDEN OF DISEASE*. A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020. Harvard University Press, US
- Murray CJ and Lopez AD (1997): Mortality by cause for eight regions of the world: Global Burden of Disease Study. *Lancet* 349: 1269–1276
- Murray CJ, Lopez AD and Jamison DT (1994): The global burden of disease in 1990: summary results, sensitivity analyses and future directions. *Bull World Health Organ* 72:495–508.
- National Cancer Center in Japan (NCC-Japan): (2004, last updated on 1st Dec.) <http://www.ncc.go.jp/en/statistics/index.html>
- National Cancer Center in Korea (NCC-Korea) (2005, copied on 15th April): <http://www.ncc.re.kr/english/>
- National Cancer Institute – Statistical Research Applications Branch (NCI-SRAB) (2005, copied on 15th April): Joinpoint software. <http://srab.cancer.gov/joinpoint>
- National Cancer Institute (NCI-USA) (2005, copied on 15th April): Surveillance, Epidemiology, and End Results (SEER). <http://seer.cancer.gov>
- National Cancer Statistics Clearing House (NCSC) (2005, copied on 15th April): <http://www.aihw.gov.au/cancer/ncsch/index.cfm>
- National Center for Health Statistics (NCHS-USA) (2005, last reviewed on 9th March): Mortality Data from the National Vital Statistics System: <http://www.cdc.gov/nchs/about/major/dvs/mortdata.htm>
- National Center of Health Information and Statistics (NCHIS) (eds) (1999): *Annual Report for National Health Information and Statistics*, Beijing, P.C. China

- National Center of Health Information Statistics and National Office for Cancer Prevention and Control (NCHIS and NOCPC) (eds) (2003): *Cancer Risk Factor Research in China*. Chinese Union Medical Univ. Publish House, Beijing, China
- National Disease Surveillance Points System (NDSPS) (2002 version): http://www.chinadsp.com/yufang/en_news/news_list_view.asp
- National Office for Cancer Prevention and Control (NOCPC) (eds) (1988): *Chinese cancer registry probative norm*. Beijing, China
- National Office for Cancer Prevention and Control and National Center of Health Information Statistics (NOCPC and NCHIS) (eds) (2001): *Cancer incidence and mortality in sample cities and counties (1988–1992) in China*. Chinese Medical Technique Publish House, Beijing, China
- National Office for Cancer Prevention and Control and National Center of Health Information Statistics (NOCPC and NCHIS) (eds) (2002): *Cancer incidence and mortality in sample cities and counties (1993–1997) in China*. Chinese Medical Technique Publish House, Beijing, China
- National Office for Cancer Prevention and Control of the Chinese Ministry of Health (NOCPC) (eds) (1979): *Investigate for Malignant Tumor Mortality in China*. People's Health Publishing House, Beijing, China
- National Office for Cancer Prevention and Control, National Center of Health Information Statistics and National Center of Cancer Registries (NOCPC, NCHIS and NCCR) (eds) (2004): *Guideline for Chinese Cancer Registration*. Chinese Union Medical Univ. Publish House, Beijing, China
- National Sexually Transmitted Disease Surveillance Cooperative Group (NSTDSCG) (1998): *Preliminary sexually transmitted diseases epidemiological analysis at national sentinel surveillance sites (1993–1996)*. *Zhongguo Yi Xue Ke Xue Yuan Xue Bao* 20: 407–412
- Niu SR, Yang GH, Chen ZM, Wang JL, Wang GH, He XZ, Schoepff H, Boreham J, Pan HC and Peto R (1998): *Emerging tobacco hazards in China: 2. Early mortality results from a prospective study*. *Br Med J* 317: 1423–1424
- North American Association of Central Cancer Registries (NAACCR) (2005, copied on 15th April): <http://www.naacr.org/>
- Nystrom L, Andersson I, Bjurstam N, Frisell J, Nordenskjold B and Rutqvist LE (2002): *Long-term effects of mammography screening: updated overview of the Swedish randomised trials*. *Lancet* 359:909–919
- Osmond C (1985): *Using age-period-cohort models to estimate future mortality rates*. *Int J Epidemiol* 14: 124–129
- Pan American Health Organization and World Health Organization (PAHO and WHO) (eds) (1995): *Health statistics from the Americas*. 1995 edition. Washington. DC, USA
- Parkin DM (2001): *Global cancer statistics in the year 2000*. *The Lancet Oncology* 2: 533–543
- Parkin DM, Bray F, Ferlay J and Pisani P. (2005a): *Global cancer statistics, 2002*. *CA Cancer J Clin*. 55: 74–108
- Parkin DM, Ferlay J, Hamdi-Cherif M, Sitas F, Thomas JO, Wabinga H and Whelan SL (eds) (2003): *Cancer in Africa*. IARC Scientific Publication No. 153. IARCPress, Lyon, France
- Parkin DM, Läärä E and Muir CS (1988): *Estimates of the worldwide frequency of sixteen major cancers in 1980*. *Int J Cancer* 41: 184–197
- Parkin DM, Muir C, Whelan SL, Gao YT, Ferlay J and Powell J (eds) (1992): *Cancer Incidence in Five Continents*. Vol VI. IARC Scientific Publications No. 120, IARCPress, Lyon, France
- Parkin DM, Pisani P and Ferlay J (1999): *Estimates of the worldwide incidence of 25 major cancers in 1990*. *Int J Cancer* 80: 827–841
- Parkin DM, Whelan SL, Ferlay J, Raymond L and Young J (eds) (1997): *Cancer Incidence in Five Continents*. Vol. VII. IARC Scientific Publications No. 143, IARCPress, Lyon, France
- Parkin DM, Whelan SL, Ferlay J and Storm H (eds) (2005b): *Cancer Incidence in Five Continents, Vol. I to VIII*. IARC CancerBase No. 7, IARCPress, Lyon, France

- Parkin DM, Whelan SL, Ferlay J, Teppo L and Thomas DB (eds) (2002): *Cancer in Five Continents Vol. VIII*. IARC Scientific Publication No.155, IARCPress, Lyon, France
- Parsons NR and Somervaille L (2000): Estimation and projection of population lung cancer trends (United Kingdom). *Cancer Causes Control* 11: 467–475
- Peng HQ, Liu SL, Mann V, Rohan T and Rawls W (1991): Human papillomavirus types 16 and 33, herpes simplex virus type 2 and other risk factors for cervical cancer in Sichuan province, China. *Int J Cancer* 47:711–716
- Percy C and Dolman A (1978): Comparison of the coding of death certificates related to cancer in seven countries. *Public Health Rep* 93:335–50
- Percy C and Muir C (1989): The international comparability of cancer mortality data: results of an international death certificate study. *Am J Epidemiol* 129:934–46.
- Percy C, Stanek E and Gloeckler L (1981): Accuracy of cancer death certificates and its effect on cancer mortality statistics. *Am J Public Health* 71: 242–250
- Pinheiro PS, Tyczynski JE, Bray F, Amado J, Matos E, Miranda AC and Limbert E (eds) (2002): *Cancer in Portugal*. IARC Technical Report No. 38. IARCPress, Lyon, France
- Pisani P, Bray F and Parkin DM (2002): Estimates of the worldwide prevalence of cancer for twenty-five sites in the adult population. *Int J Cancer* 97:72–81
- Pisani P, Parkin DM, Bray F and Ferlay J (1999): Estimates of the worldwide mortality from 25 cancers in 1990. *Int. J Cancer* 83: 18–29.
- Popkin BM (2001): The nutrition transition and obesity in the developing world. *J Nutr* 131: 871S–873S
- Population Census Office under the State Council & Department of Population, Social, Science and Technology Statistics and National Bureau of Statistic of China (PCO-NBSC) (eds) (1993): *Tabulation on the 1990 Population Census of the People’s Republic of China*. National Statistical Publishing House, Beijing, China
- Population Census Office under the State Council & Department of Population, Social, Science and Technology Statistics and National Bureau of Statistic of China (PCO-NBSC) (eds) (2002): *Tabulation on the 2000 Population Census of the People’s Republic of China*. National Statistical Publishing House, Beijing, China
- Puffer RR and Wynne-Griffith G (eds) (1967): *Patterns of Urban Mortality.: Pan-American Health Organization, Scientific Publication No. 151*. Washington DC, USA
- Qian GS, Ross RK, Yu MC, Yuan JM, Gao YT, Henderson BE, Wogan GN and Groopman JD (1994): A follow-up study of urinary markers of aflatoxin exposure and liver cancer risk in Shanghai, People’s Republic of China. *Cancer Epidemiol Biomarkers Prev.* 3: 3–10
- Qiao YL, Hou J, Yang L, He YT, Li YY, Li LD, Li SS, Lian SY and Dong ZW (2001): The trends and preventive strategies of esophageal cancer in high-risk areas of Taihang Mountains China. *Acta Acad Med Sin* 23:10–14
- Quoc NM, Hung NC and Parkin DM (1998): Cancer incidence in Ho Chi Minh City, Viet Nam, 1995–1996. *Int J Cancer* 76:472–479
- Rao C, Lopez AD, Yang G, Begg S and Ma J (2005): Evaluating national cause-of-death statistics: principles and application to the case of China. *Bull World Health Organ.* 83: 618–625
- Rao KQ, Chen YD and Tian M (1989): Research on the classification of health status in China. *Chin Health Stat.* 6: 12–17
- Rao KQ, Chen YD, Chen XZ and Mao JW (1992): Selection of a national area sample for acquisition of national health information in China. *Chin Health Statistics* 9:1–6
- Remontet L, J Esteve, AM Bouviver P Grosclaude, G Launoy, F Menegoz, C Exbrayat, B Tretare, PM Carli, AV Guizard, X Troussard, P Bercelli, M Colonna, JM Halna, G Hedelin, J Mace-Lesec’h, J Peng, A Buemi, M Velten, E Jougla, P Arveux, L Le Bodic, E Michel, M Sauvage, C Schwartz and J Faivre (2003): Cancer incidence and mortality in France over the period 1978–2000. *Rev Epidemiol Sante Publique* 51: 3–20
- Ries LAG, Eisner MP, Kosary CL, Hankey BF, Miller BA, Clegg L, Mariotto A, Feuer EJ, Edwards BK (eds) (2005): *SEER Cancer Statistics Review, 1975–2002*, National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1975_2002/, based on November 2004 SEER data submission, posted to the SEER web site 2005

- Sankaranarayanan R, Black RJ and Parkin DM (eds) (1998): *Cancer Survival in Developing Countries*. IARC Scientific Publications No. 145. IARC Press, Lyon, France
- Sen U, Sankaranarayanan R, Mandal S, Ramanakumar AV, Parkin DM and Siddiqi M (2002): Cancer patterns in eastern India: the first report of the Kolkata Cancer Registry. *Int J Cancer* 100: 86–91
- Setel PW, Sankoh O, Rao C, Velkoff VA, Mathers C, Yang G, Hemed Y, Jha P and Lopez AD (2005): Sample registration of vital events with verbal autopsy: a renewed commitment to measuring and monitoring vital statistics. *Bull World Health Organ*. 83: 611–617.
- Setiawan VW, Zhang ZF, Yu GP, Lu QY, Li YL, Lu ML, Wang MR, Guo CH, Yu SZ, Kurtz RC and Hsieh CC (2001): Protective effect of green tea on the risks of chronic gastritis and stomach cancer. *Int J Cancer* 92:600–604
- Shibuya K, Mathers CD, Boschi-Pinto C, Lopez AD and Murray CJ (2002): Global and regional estimates of cancer mortality and incidence by site: II. results for the global burden of disease 2000. *BMC Cancer* 2:37
- Shu XO, Jin F, Dai Q, Shi JR, Potter JD, Brinton LA, Hebert JR, Ruan Z, Gao YT and Zheng W (2001): Association of body size and fat distribution with risk of breast cancer among Chinese women. *Int J Cancer* 94: 449–455
- Smith KR and Mehta S (2003): The burden of disease from indoor air pollution in developing countries: comparison of estimates. *Int J Hyg Environ Health*. 206: 279–289
- State Statistical Bureau in China (SSBC) (1999): *China statistical yearbook, 1998*. State Statistical Publishing House. Beijing, China
- Stewart BW and Kleihues P (eds) (2003): *World Cancer Report*. International Agency for Research on Cancer. IARC Press, Lyon, France
- Storm H, Engholm G, Ferlay J, Langmark F, Ólafsdóttir E, Pukkala E and Talbäck M (eds) (2003): *NORDCAN: Cancer Incidence and Mortality in the Nordic Countries, Version 2.0*. Danish Cancer Society, Copenhagen, Denmark
- Sun T, Li L, Shi N and Zhang X (2003): A 40-year cohort study on cancer mortality among female workers with manual spinning of chrysotile asbestos. *Wei Sheng Yan Jiu* 32:511–513
- Sun XD, Mu R, Zhou YS, Dai XD, Qiao YL, Zhang SW, Huangfu XM, Sun J, Li LD and Lu FZ (2002): 1990–1992 mortality of stomach cancer in China. *Zhonghua Zhong Liu Za Zhi* 24: 4–8
- Sun XD, Mu R, Zhou YS, Dai XD, Zhang SW, Huangfu XM, Sun J, Li LD, Lu FZ and Qiao YL (2004): Analysis of mortality rate of stomach cancer and its trend in twenty years in China. *Zhonghua Zhong Liu Za Zhi* 26: 4–9
- Sun Z, Zhu Y, Stjernsward J, Hilleman M, Collins R, Zhen Y, Hsia CC, Lu J, Huang F, Ni Z (1991): Design and compliance of HBV vaccination trial on newborns to prevent hepatocellular carcinoma and 5-year results of its pilot study. *Cancer Detect Prev*. 15, 313–318.
- Teppo L, Hakulinen and Saxen E (1974): The prediction of cancer incidence in Finland for the year 1980 by means of cancer registry material. *Ann Clin Res* 6: 122–125
- Testa MA and Meigs JW (1980): Mathematical modeling of cancer incidence rates: Linear models of colon cancer in Connecticut. *Journal of Chronic Diseases* 33: 733–743
- The European Commission Health Promoting Programme (2005, copied on 15th April) CaMon: Comprehensive Cancer Monitoring in Europe: <http://www-dep.iarc.fr/HMP/CAMON.htm>
- Thomas DB, Gao DL, Ray RM, Wang WW, Allison CJ, Chen FL, Porter P, Hu YW, Zhao GL, Pan LD, Li W, Wu C, Coriaty Z, Evans I, Lin MG, Stalsberg H and Self SG (2002): Randomized trial of breast self-examination in Shanghai: final results. *J Natl Cancer Inst* 94:1445–1457
- Tu J (1985): Change in mortality of gastric cancer from 1963 to 1981 in Shanghai urban residents. *Zhonghua Zhong Liu Za Zhi Jul*;7: 274–276
- U.S. Preventive Services Task Force (USPSTF) (eds) (1996): *Guide to Clinical Preventive Services*, 2nd ed. International Medical Publishing, Alexandria, Virginia, USA

- United Nations, Population Division.(UN-PD) (eds) (2002): World Population Ageing 1950–2050. United Nations Publication, New York, USA
- United Nations, Population Division.(UN-PD) (2003): World Population Prospects: The 2002 Revision. Population Database. <http://esa.un.org/unpp/index.asp?panel=1>, New York, USA
- United Nations, Statistics Division (UN-SD) (2002) Economic and Social Commission for Asia and the Pacific (ESCAP): <http://www.unescap.org/stat/statdata/apinfig.htm>
- Verdecchia A, Mariotto A, Capocaccia R, Gatta G, Micheli A, Sant M and Berrino F (2001): Incidence and prevalence of all cancerous diseases in Italy: trends and implications. *Eur J Cancer* 37:1149–1157
- Villa LL, Costa RLR, Petta CA, Andrade RP, Ault KA, Giuliano AR, Wheeler CA, Koutsky LA, Malm C, Lehtinen M, Skjeldestad FE, Olsson S-E, Steinwall M, Brown DR, Kurman RJ, Ronnett BM, Stoler MH, Ferenczy A, Harper DM, Tamms GM, Yu J, Lupinacci L, Railkar R, Taddeo FJ, Jansen KU, Esser MT, Sings HL, Saah AJ and Barr E (2005): Prophylactic quadrivalent human papillomavirus (types 6,11, 16, and 18) L1 virus-like particle vaccine in young women: a randomised double-blind placebo-controlled multicentre phase II efficacy trial. *Lancet Oncology* 6:271–278
- Wang L, Lubin JH, Zhang SR, Metayer C, Xia Y, Brenner A, Shang B, Wang Z and Kleinerman RA (2000): Lung cancer and environmental tobacco smoke in a non-industrial area of China. *Int J Cancer* 88:139–145
- Wang LD, Kong LZ, Wu F, Bai YM and Burton R (2005 in press): Preventing chronic diseases in China. *Lancet* 366
- Wang JS, Huang T, Su J, Liang F, Wei Z, Liang Y, Luo H, Kuang SY, Qian GS, Sun G, He X, Kensler TW and Groopman JD (2001): Hepatocellular carcinoma and aflatoxin exposure in Zhuqing village, Fusui county, People's Republic of China. *Cancer Epidemiol Biomarkers Prev.* 10:143–146
- Wang QJ, Zhu WX and Yuan GL (1995): Trend and prediction of cancer mortality in Beijing, China, during 1980–2001. *Zhonghua Liu Xing Bing Za Zhi* 16: 195–198
- Wang RT, Wang T, Chen K, Wang JY, Zhang JP, Lin SR, Zhu YM, Zhang WM, Cao YX, Zhu CW, Yu H, Cong YJ, Zheng S and Wu BQ (2002): Helicobacter pylori infection and gastric cancer: evidence from a retrospective cohort study and nested case-control study in China. *World J Gastroenterol* 8:1103–1107
- Wang SM and Gao MY (2000): Employment and contextual impact of safe and unsafe sexual practices for STI and HIV: the situation in China. *Int J STD AIDS* 11: 536–544
- Wang TJ, Zhou BS and Shi JP (1996): Lung cancer in nonsmoking Chinese women: a case-control study. *Lung Cancer* 14 Suppl: S93–98
- Wang Z, Lubin JH and Wang L (2002): Residential radon and lung cancer risk in a high-exposure area of Gansu Province, China. *Am J Epidemiol.* 155: 554–564
- Waterhouse J, Muir C, Shanmugaratnam C and Powell J (eds) (1982): Cancer Incidence in Five Continents. Vol. IV. IARC Scientific Publications No. 42, IARC, Lyon, France
- Waterhouse J.A.H., Muir C.S., Correa P and Powell J (eds) (1976): Cancer Incidence in Five Continents. Vol. III. IARC Scientific Publications No. 15, IARC, Lyon, France
- Wiklund K, Hakulinen T and Sparén P (1992): Prediction of cancer mortality in the Nordic countries in 2005: effects of various interventions. *Eur J Can Prev.* 1: 247–258
- Wingo PA, Ries LAG, Rosenberg HN, Miller DS and Edwards BK (1998): Cancer incidence and mortality, 1973–1995. A report card for the U.S. *Cancer* 82: 1197–1207
- Wong BC, Lam SK, Wong WM, Chen JS, Zheng TT, Feng RE, Lai KC, Hu WH, Yuen ST, Leung SY, Fong DY, Ho J, Ching CK and Chen JS (2004): China Gastric Cancer Study Group. Helicobacter pylori eradication to prevent gastric cancer in a high-risk region of China: a randomized controlled trial. *JAMA* 291:187–194
- World Bank (eds) (1993): World Development Report 1993: Investing in Health, Oxford Univ Press, Oxford, UK
- World Health Organization (WHO) (eds) (1993): International Classification of Disease and Related Health Problems: Tenth Revision. World Health Organization, Geneva, Switzerland

- World Health Organization (WHO) (eds) (1994): World Health Statistics Annual 1993. World Health Organization, Geneva, Switzerland
- World Health Organization (WHO) (eds) (1997): The world health report 1997 / conquering suffering, enriching humanity. World Health Organization, Geneva, Switzerland
- World Health Organization (WHO) (eds) (1998): World Health Statistics Annual 1996. World Health Organization, Geneva, Switzerland
- World Health Organization (WHO) (eds) (2000): Health information systems development and strengthening. Guidance on needs assessment for national health information systems development. World Health Organization, Geneva, Switzerland
- World Health Organization (WHO) (eds) (2002a): National cancer control programmes: policies and managerial guidelines. Second version. World Health Organization, Geneva, Switzerland
- World Health Organization (WHO) (eds) (2002b): Immunization, Vaccines and Biologicals. http://www.who.int/vaccines-diseases/diseases/hepatitis_b.shtml. World Health Organization, Geneva, Switzerland
- World Health Organization (WHO) (eds) (2003): STEPS: A framework for surveillance. The WHO STEPwise approach to Surveillance of noncommunicable diseases (STEPS). World Health Organization, Geneva, Switzerland
- World Health Organization (WHO) (2005a, copied on 15th April): Building Health Communities and Populations: http://www.wpro.who.int/about/key_policy_orientation/policy_04.htm
- World Health Organization (WHO) (2005b, copied on 15th April): WHO Mortality Database as is WHO Statistical Information System (WHOSIS): <http://www-depdb.iarc.fr/> and <http://www3.who.int/whosis/menu.cfm?path=whosis,mort&language=english>
- World Health Organization (WHO) (2005c, copied on 15th April): Noncommunicable Disease Surveillance: http://www.who.int/ncd_surveillance/en/
- World Health Organization (WHO) (2005d, copied on 15th April): WHO SuRF Reports. http://www.who.int/ncd_surveillance/infobase/web/SuRFReports/en/index.aspx
- World Health Organization (WHO) (2005e, copied on 23th Sept.): Burden of Disease Project. <http://www.who.int/healthinfo/bodproject/en/index.html>
- Wu AH, Yang D and Pike MC (2000): A meta-analysis of soyfoods and risk of stomach cancer: the problem of potential confounders. *Cancer Epidemiol Biomarkers Prev* 9(10):1051–1058
- Wu AR (1997): Secondary prevention of cervical cancer. *Bull Chin Can* 6:8–11
- Wu-Williams AH, Dai XD, Blot W, Xu ZY, Sun XW, Xiao HP, Stone BJ, Yu SF, Feng YP and Ershow AG (1990): Lung cancer among women in north-east China. *Br J Cancer* 62:982–987
- Wu-Williams AH, Xu ZY, Blot WJ, Dai XD, Louie R, Xiao HP, Stone BJ, Sun XW, Yu SF and Feng YP (1993): Occupation and lung cancer risk among women in northern China. *Am J Ind Med*. 24: 67–79
- Xu ZY, Brown L, Pan GW, Li G, Feng YP, Guan DX, Liu TF, Liu LM, Chao RM, Sheng JH and Gao GC (1996): Lifestyle, environmental pollution and lung cancer in cities of Liaoning in northeastern China. *Lung Cancer* 14 Suppl: S149–160
- Yamaguchi N, Mizuno S, Akiba S, Sobue T and Watanabe S (1992): A 50-year projection of lung cancer deaths among Japanese males and potential impact evaluation of anti-smoking measures and screening using a computerized simulation model. *Jpn.J.Cancer Res.* 83:251–257.
- Yang CS (2000): Vitamin nutrition and gastroesophageal cancer. *J Nutr* 130(2S Suppl): 338s–339s
- Yang G, Fan L, Tan J, Qi G, Zhang Y, Samet JM, Taylor CE, Becker K and Xu J (1999): Smoking in China: findings of the 1996 National Prevalence Survey. *JAMA* 282:1247–1253
- Yang GH (1992): Selection of DSP points in second stage and their presentation. *Zhonghua Liu Xing Bing Xue Za Zhi* 13:197–201

- Yang GH (2004): China set up a Chinese population mortality and related risk factors database. *Chi J Epidemiol* 25: 642–643
- Yang GH, Huang ZJ and Tan J (1996): Priorities of disease control in China – analysis on mortality data of national disease surveillance points system. *Zhonghua Liu Xing Bing Xue Za Zhi* 17:199–202
- Yang GH, Hu JP, Rao KQ, Ma JM, Rao C and Lopez AD (2005): Mortality registration and surveillance in China: History, current situation and challenges. *Population Health Metrics* 3: 3
- Yang L, Huangpu XM, Zhang SW, Lu FZ, Sun XD, Sun J, Mu R, Li LD and Qiao YL (2003a): Changes of mortality rate for cervical cancer during 1970's and 1990's periods in China. *Zhongguo Yi Xue Ke Xue Yuan Xue Bao*. 25: 386–390.
- Yang L, Parkin D M, Li LD and Chen YD (2003b): Sources of Information on the Burden of Cancer in China. *Asian Pac J Cancer Prev*. 4:23–30
- Yokokawa Y, Ohta S, Hou J, Zhang XL, Li SS, Ping YM and Nakajima T (1999): Ecological study on the risks of esophageal cancer in Ci-Xian, China: the importance of nutritional status and the use of well water. *Int J Cancer* 83:620–624
- You W, Jin F, Gridley G, Schatzkin A, Yang G, Rosenberg P, Xiang Y, Hu Y and Li Q (2002) Trends in colorectal cancer rates in urban shanghai, 1972–1996, in relation to dietary changes. *Ann Epidemiol* 10: 469
- You WC, Zhang L, Gail MH, Chang YS, Liu WD, Ma JL, Li JY, Jin ML, Hu YR, Yang CS, Blaser MJ, Correa P, Blot WJ, Fraumeni JF Jr and Xu GW (2000): Gastric dysplasia and gastric cancer: Helicobacter pylori, serum vitamin C, and other risk factors. *J Natl Cancer Inst*. 92: 1607–1612
- Zhang B, Li X and Nakama H (2002): A case-control study on risk of changing food consumption for colorectal cancer. *Cancer Invest*. 20: 458–463
- Zhang J and Smith KR (2003): Indoor air pollution: a global health concern. *Br Med Bull*. 68: 209–225
- Zhang ZF, Parkin DM, Yu SZ, Esteve J and Yang XZ (1989): Risk factors for cancer of the cervix in a rural Chinese population. *Int J Cancer* 43:762–767
- Zheng S, Fan R and Wu Z (1997): Studies on relationship between passive smoking and lung cancer in non-smoking women *Zhonghua Yu Fang Yi Xue Za Zhi*. 31:163–165
- Zhong L, Goldberg MS, Gao YT and Jin F (1999a): Lung cancer and indoor air pollution arising from Chinese-style cooking among nonsmoking women living in Shanghai, China. *Epidemiology* 10:488–494
- Zhong L, Goldberg MS, Gao YT and Jin F (1999b): A case-control study of lung cancer and environmental tobacco smoke among nonsmoking women living in Shanghai, China. *Cancer Causes Control* 10:607–616
- Zhou BS, Wang TJ, Guan P and Wu JM (2000): Indoor air pollution and pulmonary adenocarcinoma among females: a case-control study in Shenyang, China. *Oncol Rep*. 7: 1253–1259
- Zhou YS, Zhang SW and Li LD (1997): Analysis for death for all and malignant tumors for Chinese population. *Bull Chin Can*. 6: 9–11
- Zhuo XG and Watanabe S (1999): Factor analysis of digestive cancer mortality and food consumption in 65 Chinese counties. *J Epidemiol* 9: 275–284.

13. Original Publications