Université de Montréal

Worldwide variations in sex ratio of cancer incidence: temporal and geographic patterns

Par

Syed Ahsan Raza

Département de médecine sociale et préventive
École de santé publique

Thèse présentée à la Faculté des Études Supérieures
en vue de l'obtention du grade de PhD en Santé Publique,

Option Épidémiologie

April, 2017

Résumé

Contexte: Les comparaisons internationales de taux d'incidence du cancer sont des sources importantes d'éléments pouvant aider à générer des hypothèses en lien à l'étiologie du cancer. Les estimations de la variation géographique de l'incidence de cancer peuvent être compromises par des artefacts tels que l'inexactitude et le manque de données complètes portant sur l'incidence du cancer, parmi plusieurs autres. Ces artefacts associés aux taux d'incidence pourront mener à des erreurs au niveau de l'interprétation et de la comparaison des tendances à travers les registres de cancers. Le ratio des sexes (défini comme le rapport du taux d'incidence de cancer masculin divisé par le taux d'incidence féminin) est une mesure qui pourra être moins susceptible d'avoir des ambiguïtés d'interprétation suite à de tels artefacts, dans la mesure où la mesure des cas de cancer est similaire chez les hommes et les femmes.

Objectifs: L'objectif principal de cette étude sera donc de conclure quand aux causes qui pourront générer la variabilité dans le ratio des sexes pour des types de cancers spécifiques, à travers temps et lieu, en générant des hypothèses. L'objectif secondaire sera d'explorer la mesure dans laquelle les inégalités de genre entre les pays peuvent fournir des indices sur la qualité des registres de cancer pour les types de cancer sélectionnés, à l'aide du rapport des sexes.

Méthodes: L'incidence du cancer dans cinq continents (CI-5), une base de données de registres populationnels de cancer obtenue du Centre international de recherche sur le cancer (CIRC) de l'Organisation mondiale de la santé (OMS), a été utilisée afin d'accéder aux données d'incidence de 30 différents cancers durant 3 périodes de temps (c'est-à-dire 1974-77, 1988-92 et 2003-07) provenant de 77, 142 et 281 registres sur le cancer. Des méthodes

descriptives ont été utilisées, soit des modèles à effets mixtes, pour l'étude des tendances temporelles et des variations géographiques au niveau du ratio des sexes. Afin d'explorer le biais en lien au genre, les ratios des sexes pour les cancers du *poumon*, de la *vessie*, de *l'œsophage* et du *larynx* ont été mis en concordance avec deux indices statistiques, à savoir l'Indice d'inégalité des genres de l'ONU et les estimations mondiales de prévalence du tabagisme de l'OMS.

Résultats: Les résultats obtenus à l'aide de modèles à effets mixtes utilisant un nombre égal (soit 76) registres de cancer de longue durée pour chaque année entre 1983 et 2007, après avoir ajusté le ratio des des sexes pour la variation géographique, ont démontré que le cancer du *poumon* avait le plus haut ratio des sexes en moyenne lors de la première année («baseline») (soit 9.9), suivi de *l'œsophage* (7.8), la *vessie* (5.1), le *foie* (3.8), le *pancréas* (2.1), le *rein* (1.9), la *leucémie* (1.8), le *lymphome non hodgkinien* (1.8), le *cerveau* (1.6), le *rectum et l'anus* (1.5), le *côlon* (1.2), les *yeux* (1.2), le *mélanome de la peau* (0.9), la *vésicule biliaire* (0.6) et la *thyroïde* (0.5). Dans les registres de pays ayant une faible inégalité entre les sexes et une prévalence de tabagisme similaire chez les femmes et les hommes (la Suède, la Norvège et le Danemark), le ratio des sexes pour le cancer du *poumon* était relativement bas (1.2, 1.3 et 1.6). D'un autre côté l'Espagne, tout en ayant une prévalence similaire de tabagisme chez les hommes et les femmes, montrait un ratio des sexes inhabituellement haut pour le poumon (7.1) ainsi que pour d'autres cancers associés au tabagisme (*vessie*: 14.9, *œsophage*: 10.7, *larynx*: 28.2). Les résultats de cette étude tendent à mettre en relief plusieurs types de cancer, notamment celui des *reins* pour lequel les facteurs de risques connus seront peu susceptibles de pouvoir expliquer pleinement le ratio masculin-féminin de presque 2:1, uniformément stable à travers le temps et les régions.

Conclusions: Les facteurs de risque établis dans la littérature dont la prévalence varie dans les deux sexes au niveau mondial, ne semblent pas pouvoir expliquer la stabilité du ratios des sexes pour le cancer du *rein* au cours des trois décennies. Suite à cette observation, nous avons émis l'hypothèse de certains facteurs endogènes, tels que la génétique ou la variance génétique, pouvant être en mesure d'expliquer la stabilité du ratio des sexes pour ce cancer. Un autre type de cancer, le *myélome multiple*, s'est lui aussi avéré stable à travers le temps et l'espace (le rôle de la vitamine D a été postulé). Notre étude nous a permis d'identifier des lacunes au niveau de la compréhension des causes de cancer au sein des populations.

Mots clés: *Cancer; Registres; Incidence; Régression; Modèles d'effets mixtes; Surveillance; Autocorrélations spatiales*

Summary

Context: International comparisons of cancer incidence rates are important sources of evidence for generating hypotheses about cancer etiology. The estimates of geographic variation in cancer incidence can be compromised by artifacts such as imperfect accuracy and completeness of available cancer incidence data among several others. The artifacts associated with incidence rates, can be potentially misleading when interpreting and comparing trends across cancer registries. The Sex Ratio (defined as the male-to-female cancer incidence rate) is one measure that can be less susceptible to ambiguity of interpretations by these artifacts, provided that the ascertainment of cancer cases is similar in males and females.

Objectives: Hence, the main aim of this study is to infer as to potential causes that drive sex ratio variability (i.e., the ratio of male to female incidence rates), of type specific cancers across time and geography, generating hypotheses. The secondary aim is to explore the extent to which country-level gender inequalities can provide clues on quality of cancer registries for selected cancer types through sex ratios.

Methods: Cancer Incidence in Five Continents (CI-5), a database of population-based cancer registries obtained from International Agency for Research on Cancer (IARC), was used to access incidence data on 30 different cancers in 3 time-periods (i.e., 1974-77; 1988-92 and 2003-07) from 77, 142 and 281 cancer registries. Descriptive methods were used with recourse to mixed-effect regression methods for studying temporal trends and geographic variations in sex ratios. To explore gender bias, sex ratios for cancers of *lung*, *bladder*, *esophagus*, and *larynx* were tallied with two statistics namely UN's Gender Inequality Index and WHO's global tobacco prevalence estimates.

Results: In the mixed-effect regression analysis using equal number of 76 long-standing cancer registries in each year from 1983 to 2007, and after adjusting for geographic variation in sex ratio, *lung* cancer had the highest sex ratio on average in the baseline year (i.e., 9.9), followed by *esophagus* (7.8), *bladder* (5.1), *liver* (3.8), *pancreas* (2.1), *kidney* (1.9) *leukemia* (1.8), *non-Hodgkin's lymphoma* (1.8), *brain* (1.6), *rectum and anus* (1.5), *colon* (1.2), *eye* (1.2), *melanoma of skin* (0.9), *gallbladder* (0.6), and *thyroid* (0.5). In registries belonging to countries, with low gender inequality and similar smoking prevalence in men and women (Sweden, Norway and Denmark), the sex ratio for *lung* cancer was relatively very low (1.2, 1.3 and 1.6). Whereas Spain with similar prevalence of smoking in men and women, showed an unusually high sex ratio for lung (7.1) as well as for other smoking associated cancers (*bladder*: 14.9; *esophagus*: 10.7; and *larynx*: 28.2). The results of our study also highlight several cancer types, in particular, *kidney* for which acknowledged and well-known risk factors are unlikely to fully explain the consistently stable male-female ratio of almost 2:1 across time and regions.

Conclusions: The well-established risk factors in literature whose prevalence varies worldwide in both sexes, does not seem to decipher the curiously stable sex ratios in cancer of *kidney* maintained over three decades. This observation has made us to tentatively hypothesize that some endogenous factor such as a gene or gene variant might be able to explain the stable sex ratio of this cancer. Another cancer type, *multiple myeloma* is also consistently stable across time and place, and where the role of vitamin D has previously been postulated. The study points towards gaps in our understanding of causes of cancer risk in populations.

Key words: *Cancer; Sex Ratios; Worldwide; Registries; Incidence; Regression; Mixed Effect Models; Surveillance; Spatial Autocorrelations*

Table of contents

Presentations of thesis work

1. **2017 Annual Meeting of American College of Epidemiology (ACE)**

   September 24-26, 2017

   New Orleans, Louisiana, United States.

   Poster Presentation: Worldwide sex ratio variations in cancer incidence

   Best Poster Award with cash prize

   Abstract published in *Annals of Epidemiology*

2. **2017 Global Health Conference of American Academy of Family Physicians (AAFP)**,

   October 4-6, 2017

   Houston, Texas, United States.

   Poster Presentation: Exploring the global impact of gender biases in cancer registration on reported gender disparities in cancer incidence

3. **11th International Symposium on Geospatial Health, International Society of Geospatial Health (GnosisGIS)**

   November 3-5, 2017
   Baltimore, Maryland, United States.

   Oral Presentation: Using geospatial distribution of age-adjusted gender disparities in cancer incidence to explore etiologic hypotheses

List of Figures

List of Tables

List of Appendices

List of Abbreviations

ASIR	Age-standardized Incidence Rate
CI-5	Cancer Incidence in Five Continents
DCO %	Proportion of Death Certificate Only cases
GII	Gender Inequality Index
IACR	International Association of Cancer Registries
IARC	International Agency for Research on Cancer
ICD-O	International Classification of Disease - Oncology
Im	Magnitude of Incidence
M:I	Mortality to Incidence Ratio
MV %	Proportion of cancer cases morphologically verified
RI	Random Intercept Model
RI & RS	Random Intercept & Random Slope Model
SR	Sex Ratio
SRm	Magnitude of Sex Ratio
SRv	Variance of Sex ratio
UICC	Union for International Cancer Control (Formerly International Union against Cancer)
UN	United Nations
WHO	World Health Organization

Dedicated to

To my father, who saw it all happen as a physician of more than five decades, the menace of tuberculosis, never turned his patients down no matter how poor they were, and most of all never let his family down!

Acknowledgements

"I took the one less traveled by"

Robert frost's narrative poem well illustrates my life as the one who desires to pursue life with passion and positive energy. What a life it has been on the road less travelled. Despite staying positive in life, sometimes, somewhere, you come across road blocks that makes you re-think your decisions. But pursuing PhD was never one of those decisions that I had to rethink. It was the decision that I made long time ago when I noticed that there are no qualified Epidemiologists in my native country, Pakistan. The decision to take off my white coat and stethoscope, for something that involved playing with numbers and making sense out of them was never something that I planned…. When I heard and read about the ancient folklore of *"Three Princes of Serendip"*, I realized that Epidemiology was indeed a serendipitous finding!

All the hardships that I faced and sacrifices that I made for these three letters P…H….D, are ingrained in these pages. From my childhood in Tripoli, medical school in Karachi, cancer research training in Lyon, leading to my doctoral studies in Montreal, and now a fellowship in Houston, I have met so many people that have made these hardships way less enduring. The cities and the people I met in those cities made me who I am today. Anwar Siddiqui in Karachi, Bilal Avan in London, Silvia Franceschi and Gary Clifford in Lyon, Ramzan Tahir and Jack Siemiatycki in Montreal are among the few names that I have closely worked with and have learned a lot from them. !

To my family in Karachi, words are not enough to express my love for them. My parents Zahid Raza and Yasmin Khan, who only saw me once during this whole process of doctoral

studies in Montreal are my idea of not only picture-perfect married couple but the ones who stand by all kinds of life-long hardships through all kinds of hail and storm. From my father, I inherited extra-ordinary amount of patience, and from my mother an extra-ordinary amount of silent inquisitiveness and doing good to others without claiming any credit. My brother Nauman who learned to instill the love of Parasitology among his students has a heart of pure gold, and my sister Sarah for doing an amazing job of raising two beautiful kids, Abdullah and Nasrullah who were just five and three when I left Karachi for Montreal in 2010. Sometimes, in my dreams, I can still hear them shout "Mamoo (uncle in urdu)" outside my room, to let them come inside so they can listen to World Cup cricket song by a band called "The Strings"…..*Hai Koi Hum Jaisaa? (Is there anyone like us?)*.

My thesis would not have reached to this stage, had it not been the staunch support of my wife Farrah and my in-laws in Houston. She surpassed even me in having the amount of patience that I always take pride in. Although she kept telling me many times that I can complete writing my thesis in Houston, and I stubbornly kept telling her that I need to be near the source of where everything is happening i.e., in Montreal. Little did she know that I found a little coffee shop in some obscure corner of Montreal's downtown where I would go and pounce my head around those abstract concepts in thesis that kept me awake at night.

I would like to express my sincere gratitude to Prof. Jack Siemiatycki, my doctoral advisor for the support of my Ph.D study in his research group, for his patience, guidance, and knowledge of Epidemiology. Also to Mireille Schnitzer for looking over my statistical analysis in the thesis. To Professors Maria Victoria Zunzunegui and Jennifer O'Loughlin for making me welcome in my first few months in the university. Beatrice Lauzon, my PhD colleague and office mate at CRCHUM who never hesitated in helping me out in anything

and everything. No doubt it was an added advantage in sharing office with her all these years…. translating my emails and documents to French! Thank you for everything Bea!...

To my lifelong mentor and closest friend Bilal Avan, I have come a long way with him. I have shared so many things in life with him (besides our scientific publications). He is the only person who is available all the time when I call him, whether he is in jungles of Africa or in some village in India, he would make sure to always call me. One time he had to walk a long distance in Somalia just to call me and advise me about my project. He is truly an amazing person. I am looking forward to see his son Daniel grow up into an amazing person like his dad.

My younger compatriot and partner in crime Ramzan Tahir, who hosted me in his apartment for long bouts of my visits to Montreal. Like a typical gentleman from Pakistan, he would not let me sleep on sofa at his place but will reserve it for himself. What are the odds that this young man from Sargodha happens to be a PhD in Biostatistics? No doubt, he knows all the ins and outs of my thesis. Our everyday meeting at lunch time and discussion on my thesis kept me motivated. His Punjabi jokes kept me high during bouts of depression in Montreal winter. This thesis would never have reached its final phase without him.

Finally, many thanks to Sally Campbell for her amazing capacity to retrieve papers as far back as 1871, and that too from the basement of McGill university's gym, and Lesley Richardson for listening to all my complains. And to Javier Pintos, Igor Karp, Anita Koushik, and Jerome Lavoie for all their theories during happy hours in La Reservoir! And to patience and encouragement shown by Drs. Roger Zoorob and Jason Salemi at Baylor College of Medicine, in Houston, to allow me to finish this thesis. Cheers to you all!

Statement of financial support

The acknowledgement would be incomplete without the mention of several sources of funding for my doctoral studies at Université de Montréal. In my first year of the program, I was supported by funding from *Guzzo Chair in Environmental Cancer Research* as well as *Bourse d'Excellence du PhD en Santé Publique, Université de Montréal*. I also received a supplementary award for doctoral students from the *Centre de Recherché du Centre Hospitalier de l'Université de Montréal – CR-CHUM*.

I was also supported for three years by a competitive Doctoral Research Scholarship from the *Fonds de recherche Santé Québec - FRQS, Government of Québec, Canada.*

Chapter 1 Introduction

1.1 Statement of problem: inferences from comparisons of cancer incidence trends

Cancer is a leading cause of death in many developed countries and is set to become a major cause of not only mortality but also morbidity in every region of world regardless of the country's level of resources (1, 2). Drawing on the principle of epidemiological transition related to aging, changing lifestyles and economic factors, there will be a dramatic world-wide increase in the number of cancers in the next few decades (1). It has been predicted that the numbers of incident cases worldwide will increase from 12.7 million cases in 2008 to 20.3 million cancer cases by 2030 (3). Such a transition most prominently reflects an upsurge in non-communicable diseases, of which cancer is the largest component. These changing patterns of cancer occurrence are also contingent on changes in regional distributions of known and unknown risk factors. The global burden of cancer and extent of its transition can be best described by studying differences in incidence and mortality rates in different regions. Comparisons of region-specific incidence and mortality rates of cancer aid in planning and prioritizing cancer control resources as well as facilitate monitoring and public health surveillance. Population-based studies on geographic and temporal trends of cancer are crucial in the implementation and evaluation of strategies aimed at all three levels of primary, secondary, and tertiary prevention (1, 4).

In epidemiological research, inferences based on geographic comparisons and temporal investigations of cancer incidence and mortality are complex because of some limitations and errors (4, 5). Previously, Bray and colleagues studied geographical and temporal trends of cancer in the context of epidemiological and preventive research and outlined complexities that obscured interpretation at the level of data collection, analysis and reporting results (6-

9). Investigators who intend to interpret incidence and mortality trends must therefore be aware of the characteristics of *sources* of both measures of cancer burden e.g., available incidence and mortality databases. They must be aware that certain artifacts in addition to specific interventions of interest or changes in the population prevalence of etiological factors may have impacted the trends (10). These artifacts associated with incidence rates potentially affect interpretation of trends (10, 11). These include 1) misclassification of a cancer case as a resident or non-resident; 2) duplicate registrations; 3) incorrect definition of an incident case of cancer; 4) a failure to identify or diagnose true cancer cases; 5) poor specification of diagnosis; 6) improvements in diagnostic procedures and 7) difficulties in estimating populations at the national or sub-national level, 8) and more importantly incomplete or imperfect cancer case ascertainment that may or may not be cancer type-specific or maybe the result of resources assigned to the cancer registries. On the other hand, mortality data is prone to erroneous death certification and changes in coding practices over time. If mortality trends are used as a proxy for incidence rates, further bias is introduced for cancers where prognosis has improved with time, given that case fatality would not be constant (12, 13).

The increasing availability of incidence data from cancer registries has been a major driving force in demonstrating the validity of range of analytical techniques. These include 1) graphical inspections of cancer rates that emphasize variations according to age and calendar time; 2) statistical models that augment visual graphical approaches in cancer rates; and 3) sophisticated methodological approaches such as age-period-cohort modelling that highlight generation-specific analyses (6, 10, 14, 15). In addition to the development of analytical methodologies by researchers, many cancer registries made efforts in standardizing

registration procedures that were helpful in establishing the quality and comparability of cancer incidence data (16, 17). However much of these standardization efforts have remained concentrated to registries that are well funded and where registration is part of government policy. Here, it should also be borne in mind that the history of registration in both Asia and Europe is, in fact, a history of highly motivated individuals who persuade those around them to help establish a registry. In majority of cases, it is only in very late stages that registration becomes a government concern (18). Due in large part to the efforts of the International Association for Cancer Registries (IACR) and the support of the International Agency for Research on Cancer (IARC), there are many cancer registries in Asia now generating incidence data, although the indices of quality still remain highly variable (18). One of the most important quality concern (i.e., completeness of cancer case ascertainment) is discussed in detail in Chapter 4. In Europe, cancer registries may cover national populations (e.g. Sweden, Denmark and Norway) or certain regions within a country (e.g. Italy, Spain, France, and Germany). The founding of European registries has also been a rather selective process, dependent on official policy to support and fund such activities, or through individual initiatives by research orientated clinicians and pathologists (19). As a result, European cancer registries also differ enormously with respect to the size of the population covered, the number of years of complete data available since the start of the registry, as well as, in the case of local regional registries, their representativeness of the national profile of cancer burden and risk patterns (4). Despite the limitations, cancer registries throughout history have provided useful information on the cancer burden, and have played a role in generating hypotheses about cancer etiology (20-22). The historical perspective of cancer registration and evolution of cancer registries in generating incidence data is provided in Chapter 3.

Currently, there are about 600 cancer registries operating around the world and majority (85%) are part of IACR but only those that fulfil certain quality criteria of completeness as assessed by IACR's technical staff, are then selected to be included in Cancer Incidence in Five Continents (CI-5), a joint collaborative effort of both IACR and IARC. For the last five decades, CI-5 has been the most authoritative reference on the incidence of cancer worldwide. CI-5 represents a compilation of participating cancer registries that meet certain quality criteria of accuracy and completeness (23-25). It has now been published in ten consecutive volumes since the 1960s, at approximately 5-year intervals. In the introduction to the first volume of the CI-5, Sir Richard Doll discussed the role of comparisons of cancer incidence between different regions and over time, in developing knowledge about the causes of cancer (26). They concluded that among statistics available for studying cancer, *"the most valuable data are, undoubtedly, the rates obtained by recording the occurrence of every case of cancer over a specified period."* This remains the basic function of any population-based cancer registry, which in the words of Jensen & Storm is defined as one that *"records all new cases of cancer in a defined population (most frequently a geographical area/region)"* (21, 27). Therefore, by recording every cancer case, the main goal of a registry is to provide incidence data that accurately represents the true incidence rate of cancer in the region. In practice, how closely these observed rates reflect true incidence rates, can be influenced by certain quality issues specific to the registration process in the region in which registry is situated.

The effectiveness of registries relies profoundly on the quality control procedures, which can be broken down into three components: comparability, validity, and completeness of the incidence data (16, 17). Comparability refers to the standardization of practices concerning

the classification and coding of new cases, and to the definition of incidence, such as rules for coding multiple primaries and incidental diagnoses. Validity or accuracy refers to the proportion of cases in the registry with a given characteristic that truly have that attribute, and depends on the precision of source documents and the level of expertise in abstracting, coding, and recording. Finally, completeness is the extent to which all the incident cancers occurring in a target population are included in the registry database (16, 17, 28).

There are numerous techniques used to evaluate registry completeness, the details of which are discussed later in Chapter 4. Ascertainment of cancer cases and their completeness in a registry, play an extremely important role in ensuring that the observed incidence rates are in fact true rates or at least close to them. However, not all registries can provide 100% true incidence rates, and there are not many straightforward methods that can provide accurate assessment of the extent in which all eligible cancer cases are ensured to be registered. Moreover there are cancer registries that use different case-ascertainment methods, and their completeness depends largely on the availability of local resources. Specifically, more sophisticated methods are used in registries that are well funded and less sophisticated methods are used in registries from low resource regions, and this in turn could increase the gap in completeness of data between cancer registries.

Other artifacts in interpreting incidence trends over time from cancer registries have been addressed by Saxem (29), Esteve (11), Muir (30), Swerdlow (31), and very recently by Bray (10). The required conditions that ensure truly valid comparisons of cancer trends, as described by Muir et al (30) [quoted by Bray (10)], are worth repeating unedited: 1) the definition and content of the cancer site being studied have not changed; 2) The criteria of malignancy have not changed; 3) the likelihood that a cancer will (ever) be diagnosed has not

changed; 4) the progress of cancer from inception to diagnosis is not modified by early detection or screening programmes; 5) ascertainment of incident cases and deaths has been equally efficient throughout the period of study; 6) indexing in the International Classification of Diseases (ICD) has not changed; 7) accuracy and specificity of coding is consistent over time; 8) statistics are available at the level of detail required. The authors note, few, if any, databases would meet all of the above criteria.

In the next section, we introduce a possible solution [i.e., Sex-Ratio (SR)], that can circumvent some of the problems that exists in interpretations of incidence trends and their comparisons across different geographic areas. These are then discussed in detail along with literature review in Chapter 5.

1.2 Possible solution: analyses of Sex Ratio (SR) variability of cancer incidence

The Sex Ratio (SR) (i.e., the male-to-female cancer incidence rate or vice versa) is one measure that can be useful to deal with the issue of artifacts and imperfect cancer case-ascertainment of different cancer registries across the world. Recently, the proposed "Sex-Ratio Methodology" has opened new perspectives in disease epidemiology, specifically where the etiology remains undetermined or where new hypotheses are warranted, and old hypotheses can be confirmed (32-34) . In fact, SR is a robust epidemiological marker and its variability can be used for comparing data collected from different countries and regions, and where confounding effects exerted by different factors can be supposedly minimized by carrying out this novel analytical technique (33, 35, 36). The SR has also been recently used in cancer epidemiology using country-specific or worldwide cancer registries to speculate on causes of cancers (37-39).

SR can be a useful analytical tool for exploring etiology of cancers and comparisons across worldwide cancer registries, provided that the completeness (or the incompleteness) of ascertainment of cancer cases, in registries that are compared, is similar in males and females. The key issues in cancer case ascertainment, different methods of evaluations, and completeness by sex are described in Chapter 4.

Table 1-1 shows two hypothetical scenarios in five hypothetical cancer registries: (a) when ascertainment of cancer cases differs by sex, and (b) when ascertainment is similar by sex.

Table 1-1: Hypothetical scenario of ascertainment of cancer cases is in cancer registries.

(a) when ascertainment of cancer cases differs by sex

Registry	Incidence (Males)		Incidence (Females)		Sex Ratio M/F	
	Observed	***True***	***Observed***	***True***	***Observed***	***True***
A	10	20	5	10	**2**	**2**
B	5	5	3	3	**1.7**	**1.7**
C	20	30	10	10	**2**	**3**
D	5	20	5	10	**1**	**2**
E	20	20	10	15	**2**	**1.3**

(b) when ascertainment of cancer cases is similar by sex

Registry	Incidence (Males)		Incidence (Females)		Sex Ratio M/F	
	Observed	***True***	***Observed***	***True***	***Observed***	***True***
A	10	20	5	10	**2**	**2**
B	5	5	3	3	**1.7**	**1.7**
C	20	30	6.7	10	**3**	**3**
D	5	20	2.5	10	**2**	**2**
E	20	20	15	15	**1.3**	**1.3**

By examining the observed and true incidence rates in males and females in Table 1-1 (a) and by computing SR, one can note that the observed SR differ from the true SR if the

ascertainment is good in females, but not in males (Registry C); if there is under-ascertainment in both males and females that is worse in males (Registry D); and if the case ascertainment is good in males, but not in females (Registry E). In Table 1-1 (b), males and females are equally prone to the issues of imperfect case ascertainment (i.e., case ascertainment is equally good or bad for both males and females). In this scenario the observed SR can be a valid estimator of the true SR.

Tables 1-1 hence highlights that comparison of incidence rates across registries is a function of both the true variability in incidence rates, and the relative completeness of case ascertainment across registries. Therefore, it can be useful to examine not only international variability in incidence, but also the international variability in SR, which comprises comparison that can be less susceptible to bias from incomplete ascertainment than the variability in incidence.

Much like a hypothetical scenario in Table 1-1 (b), if we assume that in CI5 based cancer registries (unlike many of IACR's non-CI5 registries), ascertainment, to be similar in males and females, and as evidenced in several sources (24, 25, 40, 41), some advantages can be envisaged when using SR compared to using absolute differences in incidence rates. It can *arguably* be a less-biased statistic which is less likely to be affected, in general, by geographical variability in diagnostic techniques, preventive strategies, tumor definitions, coding practices, and other artifacts mentioned in preceding section (37, 42). Unlike reported incidence rates, the reported SR will also be robust to these artifacts when it is used to compare international variations of cancers across registries, and that could also possibly use different methods of ascertainment (these methods are detailed in Chapter 4). Moreover, SR variation in cancer incidence as estimated from registries with presumably similar

ascertainment in males and females can be used as a distinct approach to understanding causes of variations, and making useful inferences regarding etiology.

1.3 Use of gender inequality index to explore gender-bias in cancer registration

The assumption that the completeness of cancer case ascertainment is similar in males and females in CI-5 based cancer registries is explored in this dissertation through United Nation's Gender Inequality Index or (GII) on selected cancer types (specifically by tallying SR of cancers such as *lung, bladder, esophagus, and larynx* along with country-specific smoking prevalence). Since this assumption of male-female similarity, can be critiqued based on the possible existence of gender-bias (i.e., differential disparities in health seeking actions such as access to health care, diagnostic services, and treatment of cancers that might not be equal in both genders) (43, 44), a well-recognized multidimensional indicator such as gender inequality index can be used in the context of exploring gender-bias in cancer registries.

The measurement of gender inequality has received increasing attention over the past few years (45, 46) and has been explored in epidemiological studies (47, 48). The Gender Inequality Index has been designed to capture gender inequality through relatively new functional form to summarize multidimensional information into a real number that can be used to compare countries' performance in this domain over time. The Gender Inequality Index reflects gender-based disadvantage in three dimensions namely: reproductive health, empowerment and the labor market, for 160 countries. It shows the loss in potential human development due to inequality between male and female achievements in these three dimensions. It ranges from 0, where women and men fare equally, to 1, where one gender fares as poorly as possible in all measured dimensions (49). As of 2015 data, the lowest

gender inequality country is Switzerland (GII: 0.04) and the highest gender inequality of 0.77 is found in Yemen (49).

According to Permanyer, the interest in measuring gender inequalities has some instrumental motivations (45, 46). The existence of gender inequalities is related, sometimes in a complex and intertwined way, to socio-economic aspects (hence the term gender) which can be very relevant from the policy making point of view. This is explained through an example of the links between gender and fertility levels in a country. There is empirical and theoretical evidence suggesting that in the countries where gender relations are more egalitarian, the fertility levels tend to be lower (50, 51). Another reason why there is a great interest in measuring gender inequalities is its presumed link with countries' economic growth. Some policy based working papers have tested empirically whether high gender inequality levels in a given country affect its economic growth. Klasen (52) and Dollar and Gatti (53) have suggested that the higher the gender equality, the higher the growth rate.

These arguments as implied by Permnayer (45) allude that there are not only good enough reasons for using valid methods of computing gender inequality in multidimensional contexts, but also, that it could be applied in a way to stimulate inquiries in unexplored domains, to raise more attention to gender inequality and its reduction. This dissertation research uses gender inequality index to highlight potential gender-bias, if existing, in the domain of cancer registration, specifically in CI5 based cancer registries.

In summary, the principal focus of this thesis is to present SR variability through the estimation of geographic and temporal variation in the magnitude of SR of cancer incidence for different types of cancer. The underlying premise is that geographic and temporal variability in the SR are less susceptible to bias than geographic and temporal variability in

estimated incidence rates, and that these can therefore provide a window to researchers in understanding the risk factors that might explain not only geographic and temporal variability in the SR but even geographic and temporal variability in incidence of cancers. This analysis is also further extended in exploring whether there is any pattern for a cancer type when the SR is geographically clustered, and whether any inference can be made from this pattern. Furthermore, the assumption that the completeness of cancer case ascertainment is similar in males and females is also further explored in this dissertation through United Nation's Gender Inequality Index or (GII). This type of analysis can also provide clues on quality of cancer registries in CI-5, and can inform the public health debate surrounding the contextual problem of gender-bias in cancer registration.

1.4 Outline of subsequent chapters

Following on from this introductory chapter, Chapter 2 introduces two important measures of cancer burden namely incidence and mortality. Some factors that complicate assessments of incidence and mortality of certain cancer types are discussed. The relative merits of both measures are highlighted and then rationalized for thesis as to why incidence is preferred for the current analysis. The usefulness and limitations of both are appraised in depth in terms of generating and testing etiologic hypotheses and providing information for resource planning. We then describe geographic variations in cancer incidence and the usefulness of these rates in comparing different populations. Chapter 2 ends with the section on review of methods to analyze cancer trends e.g., 1) exploratory analyses through graphical presentations (that includes utility of age-standardization as well as available scales such as arithmetic or log-transformations to plot these rates over time); 2) quantification of temporal

changes through traditional models and modification of those models; and 3) a final brief note on the use of more sophisticated models.

After discussing the relative merits of incidence and the mortality for trend analysis, and justifying the advantages of incidence in Chapter 2, we then introduce the historical context of worldwide cancer registration in Chapter 3. A flurry of research activities arrived with the advent of cancer registries. Therefore, the purpose of this chapter is to apprise the reader the importance of cancer registries, the role of several individuals and organizations, and the types of existing registries in generating the incidence data.

Chapter 4 provides an overview of completeness of cancer case-ascertainment, one of the important artifact hindering valid interpretations of comparisons of incidence trends across worldwide cancer registries. This chapter highlights the different methods used to assess the completeness, and then comparisons of completeness indices of these methods in different cancer registries are reviewed. The chapter rounds off with observations on completeness of cancer case ascertainment by sex (i.e., similarity or lack thereof in males and females) with the acknowledgement of contextual obstacle in cancer registration i.e., gender-bias.

In chapter 5, literature review is carried out regarding utility of sex-ratio methodology in different diseases. Previous studies using sex ratios in incidence or mortality of different diseases such as multiple sclerosis, Parkinson's disease, rheumatoid diseases, infectious diseases, and cardiovascular diseases as a tool to generate hypotheses are highlighted in addition to cancer. The purpose of this chapter is to illustrate that sex ratios of incidence and/or mortality have been successfully used in other diseases and that the variability of SR can be a used as a guide to interpret trends and help investigators to generate hypotheses regarding cancer etiology.

Chapter 6 presents rationale and objectives of the study. Chapter 7 presents the methodology used to carryout analysis on sex ratio variability, first in three-time periods, and then followed by analysis on long standing cancer registries. Details on mixed-effects regression modelling, spatial analysis, and analysis on gender inequality are also presented. Chapter 8, presents results such as descriptive statistics, regression models and the interpretations, spatial autocorrelations through Moran's Index in two continents, and also predicted average sex ratio changes in cancer registries according to low and high gender inequality countries in four types of smoking related cancers.

Finally, in Chapter 9, we discuss the results, and highlights the value of inductive form of inferences based on observations and trends, and present tentative hypotheses and raise few questions based on observations and trends. In this chapter we also theorize the observations based on self-point of view, which is again an important part of inductive reasoning. We discuss the role of possible gender bias in cancer registration and recognize strengths and weaknesses of the study. The thesis ends with some concluding remarks and acknowledgement on the important gaps that exists in understanding causes of cancers.

Chapter 2 Background Review on measures of cancer burden

The text in chapter 2 that follows discusses two most important measures of cancer burden i.e., incidence and mortality, and their complementary and contrasting nature. These two statistics in which cancer burden is expressed are both useful in the investigation of cancer etiology (54-59) as well as for understanding cancer disparities or inequalities in different parts of the world (60).

In planning and evaluation of public health strategies, temporal investigations of cancers have important implications, yet they are complex phenomenon to study because of limitations associated with them (as pointed out in Chapter 1). Investigations of the changing temporal patterns of cancer incidence and mortality are considered standard epidemiological tools in public health surveillance. Long standing data from cancer registries and vital sources enables quantification of incidence and mortality rates over time and may provide clues as to the underlying determinants (7, 8). Changing rates over time is 'supporting evidence' to inferences regarding causality if the temporal patterns make sense, and provided that sufficient time lag is there, to the (known or unknown) distribution and prevalence of one or several risk factors. These analyses of incidence and mortality trends may generate or establish novel hypotheses or provide confirmatory evidence of existing ones.

Ecological studies based on geographic variation of cancer incidence and/or mortality rates have made important contributions to exposure-cancer hypotheses (55). For example, the hypothesis on ultraviolet-B -Vitamin D and cancer was first proposed based on the map of colon cancer mortality rates in the US which portrayed that the mortality rates were correlated with annual sunlight doses (61). The investigators proposed that because Vitamin D production was the most important physiological effect of sunlight in humans, it suggested

mechanism linking sunlight to a reduced risk of cancer. In the ensuing years, the hypothesis was extended to other types of cancers across a wide range of geographic regions (59, 62-68).

The international pattern of cancer incidence and mortality can point to regions of the world where research efforts may be particularly worthwhile e.g., comparisons of human papilloma virus infection in Denmark and Greenland with a five-fold difference in cervical cancer incidence led to preventive efforts in those countries in next decades (21, 69). International comparison of cancer occurrence therefore provides clues to etiology, and the demonstration of variation in incidence (and mortality) has made an important contribution to the recognition of the environmental origin of many cancer types (70-72). Statistics by age and sex show widely different patterns and variations between cancer types. In their textbook on cancer registration and principles, Jensen *et al*., writes that such basic features of measures of cancer burden may not always be easily understood and explained, but they should provoke the curiosity and are useful in the generation of etiological hypotheses (21).

Therefore, in many instances, it has been shown that efforts to reduce the global cancer burden, its causes, and disparities can be initiated with an understanding of geographic patterns in cancer incidence and mortality. The next sections present definitions of cancer incidence and mortality, and why incidence can be a better measure of cancer burden than mortality in terms of understanding cancer causes.

2.1 Cancer incidence

2.1.1 Definition

Cancer incidence is defined as the number of new cancer cases occurring in a defined population within a specified period of time. The number of new cancer cases is commonly expressed as a rate per 100,000 persons per year that approximates the average risk of developing a cancer, and is used for comparisons between populations (73). Age-standardized incidence rates (ASIR) are used for comparison purposes between populations that have different age structures. The term rate is often used interchangeably with the risk of developing a cancer, but, strictly speaking, risk is a proportion and describes the accumulation of the effect of rates over a given period of time e.g. the cumulative risk (40). Incidence is determined by exposure to etiologic factors and individual susceptibility and may be further affected by screening practices, health care access, and quality of care (73).

2.1.2 Sources of data and quality

Cancer incidence data are product of population-based cancer registries, whose main function is to collect and classify information on all new cases of cancer in a defined population, and provide statistics for assessing the impact of cancer in the population (21, 74). The recording of individuals with cancer followed several failed attempts at producing good quality cancer morbidity statistics. Cancer surveys in Europe in the first decade of twentieth century resulted in poor participation rates, while analyses of several metropolitan areas in the U.S. in 1937-38, 1947-48 and 1969-71 were eventually considered to be not of much use (75). However innovations in methodologies from pilot studies in the 1930s brought about a more successful system that reported cases by name, eliminating multiple

registrations and allowing the follow-up of individual patients (75). Background on the inception of cancer registration and its utility in surveillance is provided in Chapter 3.

Cancer registries are effectively utilized in incidence-trend studies provided that the quality control measures are well placed. The purpose of these registries is also to produce timely information on the incidence, and as such, play a pivotal role in public health and epidemiological research. One of the most important quality issue is the completeness of cancer case ascertainment which is the extent to which all the incident cancers occurring in a target population are included in the registry database. There are numerous techniques used to evaluate registry completeness (17), including: 1) methods that evaluate the data sources themselves (number of sources/notifications per case, percentage of cases histologically and morphologically verified (%MV), and methods based on death certificates); 2) methods that involve independent case ascertainment (rescreening of cases, capture-recapture methods, the mortality:incidence (M:I) ratio); and 3) historic data methods (stability of incidence over time, comparison of incidence in different populations, age-specific incidence curves). Some of these methods that are broadly classified as quantitative and semi-quantitative are critically reviewed in Chapter 4.

2.1.3 Factors affecting assessments of incidence over time

Concerning the issues of detectable artifacts (e.g., instances where specific artifacts for a specific cancer type can be recognized), Bray (6, 10) highlighted several seminal papers by Saxen (29, 76) and Muir (30) in his discussion on factors that complicate assessments of certain cancer types over time. These researchers have listed some of the factors as follows:

a) International Classification of Disease (ICD):

Changes in classification and codes of ICD has brought about the possibility of artefactual changes in time trends. Changes in the content of the ICD in consecutive revisions have had considerable effects on the evaluation of time trends; in particular with cancers of the *lung* and *liver*. The demand for a better delivery of detail in classification at the level of subsite in each successive volume of ICD has led to an awareness of lack of comparability.

b) Definition of malignancy:

The definition of the tumor can change over time since there has been an increasing likelihood of observing evidence of malignancy in tissue samples through improving technology by pathologists. For example, the increase in cancers of the thyroid may, at least in part, be due to an increasing tendency to interpret papillary change as malignant. Registry practices regarding the coding of invasiveness of bladder tumors have also shifted accordingly (6).

c) Latent carcinoma:

There is also an increasing likelihood of incidental diagnoses of tumors that may not have progressed to invasion. The ICD does not make any provision for such cancers, and interpretation of cancer incidence over time should, therefore, take a particularly cautious approach. For example, latent carcinomas of a particular type that are coded as malignant are influenced by some intervention or opportunistic screening. A classic example can be prostate coded as malignant which can be influenced by trans-urethral resection of prostate and prostate- specific-antigen.

d) Effects of screening programs

Stage of progression at which cancer or pre-cancerous lesions are detected are modified by screening programs. Slow-growing tumors are therefore more likely to be diagnosed than under normal conditions. For example, in the case of breast screening programs, the classical model involves a temporary, artificial, increase in the observed incidence as a result of the early diagnosis of malignancies that would have eventually become clinically manifested in time.

e) Changes in medical practice

Changes in the trends regarding presentation of patients for diagnosis, the availability of medical services and the ability of the doctor to make diagnoses can influence the likelihood of a cancer being diagnosed, as well as the accuracy of the recorded information. Examples include an increasingly aggressive investigation of illness in the elderly, e.g. for brain tumors, there is an inclination towards greater specialization in the field, increasing use of treatment guidelines, and specialist referrals.

f) Population denominators

In the definition for incidence rate, the denominator is the person-time usually taken from population estimates. These people should be at risk of having a neoplasm, however this assumption is not met in some cases. An important example is the need for adjustment for prevalence of hysterectomy in the female population in the study of uterine cancer trends e.g., Luoto *et al*., used information on hysterectomy to derive uteri-at risk and cervices-at-risk populations to correct gynaecological cancer rates in Finland (77). The premise of their study was that an increasing proportion of women in some European countries have undergone a hysterectomy in the last thirty years and are consequently are at minimal risk of

certain gynecological cancers, particularly endometrial cancer. This is the case where an adjustment to the person-time would be desirable. Unadjusted trends may not provide a more accurate temporal description that does not account for the prevalence of a condition (in this case, hysterectomy) in the population.

g) Registration practices

With time, improvement in the completeness of registration may also produce artefactual changes in the incidence trends. Registries may differ in their operational rules regarding the inclusion of cancer types where there is difficulty in distinguishing between malignant, benign and unspecified tumors. The major problems are seen in tumors of the brain and bladder, although trends in melanoma of the skin and thyroid may also be affected by this artifact. Trends in bladder cancer incidence are very difficult to interpret without precise information on how registries have dealt with papillomas over time. It has been shown that, on the exclusion of papillomas of the bladder, much of the variation in bladder cancer incidence rates in the Scandinavian countries was removed (29, 76). Overall, changes in registration practises over time are more likely to affect comparisons between registries, rather than trends in a single registry (10, 30).

2.1.4 Geographic variations in incidence rates

In 2012, there were an estimated 14 million new cancer cases worldwide, with 45% of these cases seen in Asia, 26.0% in Europe, 14.5% in North America, 7.1% in Central/South America, 6.0% in Africa, and 1.0% in Oceania (73). Cancers of the lung, stomach, colon and rectum, liver, and esophagus have the highest incidence worldwide (55, 73).

Cancer incidence varies considerably across geographic regions, and this variation is generally more pronounced between less- and more-developed regions (74). As discussed in

the preceeding sections of this chapter, the variability is cancer-specific; that is, for some types of cancer, region A may have higher rates than region B, while for other types, region B may have higher rates than region A. This variability could be due to true differences in incidence resulting from inter-regional variability in cancer risk factors, and/or it can be due to artifacts. International comparisons of cancer incidence have been fundamental in the development of hypotheses regarding cancer etiology (3, 67, 72) which are typically addressed in analytical studies. As well as providing hypotheses for investigation, inter-regional comparisons of cancer incidence also provide information to identify which populations to study to address specific hypotheses (19).

There are numerous examples illustrating that geographic comparisons of cancer incidence give rise to hypotheses that were eventually instrumental in establishing cancer risk factors. (3, 67, 72, 75-77). Lemrow *et al*. and Wiggins *et al*. analyzed cancer incidence rates in the USA and found that incidence rates varied among Native Alaskan populations and Native populations in other parts of North America, and also that they often differed from rates among non-Hispanic whites (78-80). The study hypothesized the effects of behavioral causes such as obesity, tobacco use, and physical inactivity. Epidemiological studies have been carried out to confirm the association of these factors along with infectious agents and genetic factors in cancers (78)(81-85). Micheli *et al*. reported large differences in the prevalence of certain cancers in the USA (86) and in countries included in the European registry (87). Richer areas in Europe had higher prevalence rates, suggesting that incidence varied with economic development.

In China, investigators noted a very high incidence of esophageal cancer in a small city in the northeast of the country and in towns located around the city (88). Speculation about

environmental and dietary factors were underpinnings of findings that ensued with theories on temperature of the food, scraping of the esophageal lining due to eating dried corn husks, eating moldy bread, and the quality of the soil in the region. Investigators in a subsequent case-control study also noted a novel serological association with the human papilloma virus and the risk of esophageal cancer while working on previous theories (89). This was followed by several studies carried out in other regions in China and worldwide (90-92).

Studies comparing incidence rates of cancers have historically provided significant leads about environmental, nutritional and behavioral causes. The often cited estimate that 80% of the cancer burden worldwide is due to environmental factors is derived from observed geographical variation in cancer rates (67). Comparisons of incidence rates have also been useful in migrant studies of relocated ethnic populations, with rates that dominate in the country of origin as well as in the host countries (93-95). These studies have been useful in understanding the relative importance of genetic and environmental factors in cancer etiology.

Comparisons of incidence rates worldwide serves as the background for creating international and national strategies for cancer prevention and control. High cancer incidence rates affect national economies, social development programs, and national public health system of individual countries. The information on incidence trends is useful in making plans for health promotion and serves as a guide for future scientific research. Therefore, obtaining complete and accurate information on cancer incidence is crucial (78).

As detailed earlier, the cancer incidence rate is a preferred measure for uncovering differences in the geographic occurrence of cancer. Unfortunately, in many regions, especially, but not limited to less developed regions, the health care systems and the cancer

registration resources are insufficient to provide reliable data on incidence. Among many problems, there is unequal distribution of health care resources, difficulty in gaining access to information about people living in remote rural areas of the country, and non-availability of advanced technical tools that are required for diagnosing cancer. Even when diagnostic technology is available, there is often no reliable registration system for recording cancer incidence rates in a defined population. In countries with few resources, complications arise due to imperfect cancer case ascertainment, and different lower quality methods relative to more developed regions. As such, incidence comparisons between different regions can be biased.

2.1.4.1 Within-country and between-country variability

As reviewed in this section on geographic variability of cancer, there is a rich literature available on comparisons of incidence rates across wide range of geographical regions, yet the comparisons of variability of incidence rates within countries and variability between countries is almost non-existent. Recently, Sang *et al*., have attempted to provide the first ever blue print for researchers to explore and compare the variability of cancer incidence rates within and between countries. They fitted a weighted multilevel Poisson regression models with a random country effect for different cancer types and sex separately (79). Cancer registries from CI5-vol X were used as surrogate for countries across Europe and North America for 10 invasive cancer types. Correlations of age standardized incidence rates within countries were termed as *intra-class correlation* coefficients that was labelled as *intra-country correlations*. This coefficient ranged from 0 to 1, with 0.5 indicating equal variances between and within countries (intra-class correlation of less than 0.5 indicated within-county variance to be larger than between-country variances, and value greater than

0.5 indicated within-country variance to be less than between-country variance). The four cancer types with the highest intra-class correlation (i.e., within- < between-country variability) among men included *prostate* cancer, *melanoma* of the skin, stomach and *lung* cancer. Among women, *lung* cancer, *breast* cancer, *melanoma* of the skin, and *cervical cancer* had the highest intra-class correlation. Cancer types where within- was greater than between-country variation were cancers of *pancreas* and *leukemia* in both men and women (i.e., low intra-class correlation of < 0.5). This analysis of within- and between-country variability through the use of cancer registries was able to offer etiological clues, that cancer types where within- is less than between-country variability, random mutations or number of stem cell divisions can be the likely explanations. On the other hand, lifestyle and environmental factors might play a role where between- is larger than within-country variability.

Also, recently, degree of within- and between- geographical variability was assessed through a new index measure using cancer registries from high quality Nordic registries, however no hypotheses generation exercise was undertaken from the analyses (80).

2.1.4.2 Spatial variability

According to Kelsall and Wakefield (81), a valuable public health practice is to examine incidence rates across geographical regions, and understanding spatial variation of patterns can provide insight as to their causes and controls. Spatial data refer to data with locational attributes and spatial variability occurs when a quantity that is measured at different spatial locations exhibits values that differ across the locations (82). Spatial variation in disease risk is of interest in epidemiology because it facilitates the display of incidence (and mortality) rates geographically for descriptive purposes, inform public health preventive strategies, and

is also helpful in assessing success of preventive programs such as screening. Most importantly, modelling of spatial variation helps formulate and provide evidence for hypotheses concerning disease etiology through ecological studies (81). The explosion in data gathering, linkage and analysis capabilities in recent times has been further expanded by advanced computing technologies such as geographic information systems, which in turn has greatly improved the ability to measure and assess varying spatial patterns of cancer trends. Cancer registry data has been widely used in analyzing spatial variability of specific types of cancers and assessing spatial clustering (83, 84).

Spatial variations in cancer trends are discussed in several papers, especially in the last few decades (85). The most common cancer types being *breast, lung*, and *prostate* from the prevention point of view. Some of these papers provide emphasis on *clustering* for analyzing spatial patterns, that further facilitate understanding of causes of cancers. These clusters were defined as unusual agglomeration, e.g., of high or low incidence of cancer. Most of these studies preferred using incidence rates rather than mortality for studying spatial variations since incidence rates have advantage over the former, in terms of providing information on histological characteristics of cancer, and due to ease of comparison across wider range of countries because of availability of data from cancer registries (83, 85). Moreover, survival rate of one cancer type can vary according to geographical area which can hamper the geographical comparison of mortality data. In these spatial studies as reviewed by Roquette *et al*., (85), the association between cancer incidence and possible causes were also discussed, yet the results remained largely inconclusive.

Given the vast scope of cancer research being carried out worldwide, research themes with spatial dimension are becoming crucial in terms of generating hypotheses. Spatial

analysis can be used to answer questions such as “Why do some geographic areas have high rates of certain types of cancer?”. It can be used in deductive and inductive form of reasoning (86), wherein deductive reasoning begins with thinking up of theory, that is narrowed down to more specific hypotheses that can be tested. It is then narrowed down where observation is made to address the hypotheses, and then test hypotheses, and finally confirm or not confirm the original theory. Inductive reasoning works the other way around and remains open ended and exploratory (86). These inductive and deductive applications of geospatial analyses can be used to test the validity of relationship between cause and occurrence of cancer. Research on cancer’s spatial variability, hence, represents an important resource for decision-making and policies design to fight one of the most important diseases known.

2.2 Cancer mortality:

2.2.1 Definition

This measure of the burden is defined as the number of cancer deaths occurring in a defined population within a specified period. It is influenced by cancer incidence, individual biological factors, tumor characteristics and stage at diagnosis, and response to available treatment (60). For a given type of cancer, mortality rate is a function of incidence rate and fatality rate (proportion of cancer cases that result in death).

2.2.2 Sources of data and quality

Mortality data’s relative advantage over incidence is because of its more comprehensive availability at the national level for many countries. The WHO mortality database has mortality statistics for most countries, and for more extended periods than that of incidence. Because of this reason, temporal analyses of cancer mortality might be more common than

those of incidence in the medical literature. There are, however, some potential difficulties in interpreting mortality rates for some cancers, and are discussed in this section.

As pointed out by Boyle (87), asides from artifacts related to cancer registration practices, many of the factors that affect incidence equally affect mortality data because they both rely on the accuracy of the initial cancer diagnosis. Comprehensive mortality statistics require that diagnostic data are available on those who died, which are then transferred to death certificates, and which are then accurately coded, compiled and analyzed. Death registrations require that the correct diagnosis is mentioned on the death certificate and that this diagnosis is confirmed as the underlying cause of death (87).

Previous studies, more specifically by Percy *et al*., have investigated the accuracy of death certificate diagnoses in vital statistics data e.g., by comparing cause of death entered on the death certificate with a reference diagnosis derived from autopsy reports, detailed clinical records, or cancer registry data (88-91). These studies have shown that the level of accuracy of the recorded cause of death declined as precision in the diagnosis increased. A tendency to over-record non-specific diagnoses instead of the exact site (e.g. large intestine instead of rectum) has been noted, and accuracy is often lower in older ages. A study on cancer trends in England and Wales during three decades from seventies to nineties in the age group 75-84 found the rise in all-cancer mortality in the older age group was in part due to increasing lung cancer mortality, but the study also concluded that the data artifacts were also responsible for much of the increase in the other common cancers (92). Their final conclusion was that use of routine mortality statistics lacks validity at older ages because of imprecision in certification of cause of death.

Death certificates were compared mentioning cancer as the underlying cause of death for 50,000 incident cases in a study by Percy *et al* (89). The accuracy of death certificate was assessed by comparing the primary cancer site reported on hospital diagnosis with the cancer site coded as underlying cause of death. Over-reporting on death certificates was indicated when the detection rate (proportion of hospital diagnosis with a cancer of certain type in which cause of death reflects the same hospital diagnosis) was high relative to the confirmation rate (i.e., proportion of cancer deaths for which the underlying cause was confirmed by hospital diagnosis), as was the case for cancer of larynx and colon, and unspecified cancers of the uterus. Underreporting was observed (confirmation high relative to detection rates) for cancers of the cervix and rectal cancers. Death certification reported melanoma correctly in most cases. Percy *et al*., also carried out the comparison of coding of cancer-related death certificates in seven countries (Canada, United States, France, United Kingdom, West Germany, Norway, and Soviet Union) and found marked differences between these countries in the allocation of lCD-codes to death certificate diagnoses (91). They concluded that these differences could seriously affect cancer mortality statistics.

2.3 Other measures of cancer burden

Other measures of cancer burden such as survival and prevalence have also been used as indicators by researchers and health care planners (93-96). Survival of cancer patient is defined as the time elapsed between diagnosis and death. However, it has been pointed out that survival figures are not suitable for geographic and temporal trend analyses because of underlying trends in registration practices (survival estimates are also produced by cancer registries). Survival comparisons have limitations when understanding cancer etiology, and their use is more focused on understanding treatment protocols of cancer patients. Most

notably, follow-up to confirm survival may be incomplete in many cancer registries (97). Follow-up of registered cancer cases quite often require matching death certificates against cancer notifications and assuming that unmatched cases are still alive. None-the-less, survival is useful indicator of the effectiveness of cancer treatment (93).

Prevalence is a more complex measure of incidence, fatality, and other influences operating in affected individuals prior to death or cure. Total prevalence of cancer is the number of persons in a defined population alive at a given time who have had cancer diagnosed at some point in time in past. It reflects the number of individuals in the population requiring at the same time a defined care procedure. It is not particularly useful for health care planning if the proportion of long term survivors can be considered cured. It can be informative if it reflects the number of subjects being in a given status rather than that of subjects ever diagnosed with the disease. For example, patients with diabetes and hypertension require continuous treatment but this cannot be the case for cancer at several sites, which when treatment is successful, can be considered cured (94).

2.4 Incidence versus mortality

As discussed, incidence and mortality are the two most important basic measures to monitor the risk of cancer in the population. They are the fundamental data resource for studies on cancer causation. Several researchers have deliberated on the relative merits of incidence and mortality and to time trend analyses (11, 72, 87, 97). The importance of determining artefacts and considering their contribution to observed cancer incidence and mortality trends have been addressed at length by Saxem (76) and Muir *et al* (30, 98), while Percy *et al.,* have investigated the accuracy of death certificate diagnosis (88, 89, 91)

Analyzing incidence trends may allow some insight into the possible changes in the prevalence of risk factors that drive the trend. Mortality rate, given its coverage and availability, has historically been a useful surrogate for incidence. The assumption of constancy over time in the case fatality ratio however may not hold for cancers where prognosis has been improving and novel effective therapies for a number of cancers were introduced in past. Mortality rates are certainly the useful measure of disease outcome; in determining the beneficial effects of a specific treatment regimen at the population level (8).

But in the same time period when survival was improving, the cancer incidence in some Nordic countries increased, thereby allowing detailed analyses of secular trends (30). However, in their landmark study of environmental causes of cancer, Doll and Peto preferred to utilise mortality rather than incidence data, pointing out that incidence was more complex to interpret (72). In addition to the artefacts associated with changing practices in cancer registration process, they considered the effects of changing practice in classification of cancers associated with different rates of fatality and the spread of screening tools that detect cases earlier as having had a large impact on the rising incidence of many cancer types. This neglect of available incidence data from them drew some criticism from Clemmesen and Nielsen (99) and Devesa *et al* (100) although Doll *et al* later clarified their viewpoints on mortality and incidence in their response to a letter from Clemmesen (99) by stating that *"Indeed, in many parts of the world where morbidity data are poor, cancer registry data may be far more reliable"*.

Most researchers have an agreement that a combined description of trends in incidence and mortality often serves to clarify understanding of the underlying biological processes,

and to have a balanced discussion in interpretation keeping in mind the relative strengths and weaknesses of each measure (30, 72, 87, 100).

In summary, while we do not understate the importance of mortality rates, there is a need to make a statement in dissertation as to why incidence is needed to understand cancer etiology rather than mortality. As discussed before in this chapter, since the mortality rate is partly determined by the fatality rate, it is an imperfect indicator of the incidence rate, and variability in the mortality rate is an imperfect indicator of variability of the incidence rate. Thus, if the purpose of geographic comparisons in rates is to elucidate possible etiologic factors, it is incidence rates rather than mortality rates that should be examined. For cancer types with poor survival, the mortality rate may closely approximate the incidence rate, and in such circumstances the mortality rate may serve as a useful proxy for the incidence rate (73, 101). However, for cancers of moderate to good survival, the mortality rate is a poor proxy for incidence, and variability in mortality may reflect variability in features like screening and treatment that have nothing to do with etiologic risk factors.

2.5 Review of methods to analyze cancer trends

This section presents a background review and critique of the various approaches that are available to a researcher if the intent is to analyze cancer occurrence over time. In putting forward propositions for an appropriate analysis, researchers over the years have emphasized on the need for right approaches to analysis and presentation, that can maximize comparability of trends over time, and with proper interpretations, to make sense of any previous knowledge regarding the biological process of cancer (102-107). This section of the chapter discusses the following approaches to trend analyses: 1) use of graphical descriptions

as exploratory analyses, 2) use of traditional models to quantify temporal change, and finally, 3) use of more sophisticated models such as age-period-cohort modeling techniques.

2.5.1 Exploratory analyses through graphical depictions

A graph can be used as an analogy for story-telling or rather more appropriately story-telling showing *some evidence of something*. It presents data visualization in a way that becomes a window to communicate results effectively and speculate into future. Hence, one can use these graphs as a pivotal point in one's story. According to classical book by Tufte (108) [also quoted by Devesa *et al* (103)], "Excellence in statistical graphics consists of complex ideas communicated with clarity, precision, and efficiency". Graphical approaches remain an intrinsic part of any good data analysis, despite many recent advances in statistical modelling. Over the years, a variety of graphical approaches have been used to visually portray and analyze temporal trends, especially annual rates of change, in cancer incidence and mortality (8, 30, 102, 103, 109). The next subsections underpin common and useful attributes of graphical presentations in cancer trend analyses e.g., choices regarding the form of rate (such as age-standardized) or selection of the scale (such as arithmetic or log-transformation of Y scale).

2.5.1.1 Choice of rate: age-standardized or age-stratum-specific

Age is a strong determinant of cancer risk. According to Bray and other researchers, there are very strong reasons to adjust for the effects of age when comparing cancer risk in populations over time (110). About ninety percent of human cancers are epithelial, and according to Armitage and Doll (111), these types of cancers increase approximately as a fifth power of age, and represent about a 1000-fold difference in cancer rates between young

people aged 20 and older persons aged more than 80. Also, the demographic effects of aging and population growth continue to have a major impact worldwide.

Direct standardization procedures yield the age-standardized incidence rate (ASIR) and cumulative risk, both of which absorb the chronological sequence of age-specific rates (e.g., in 5-year stratum specific age group, 5 to 9, 10 to14, 15 to19 and so on), allowing comparisons of cancer risk over time in the same population using a single summary measure (40). In other words, direct standardization yields a standardized rate, which is a weighted average of the age-specific rates, for each of the populations to be compared. In ASIR, a standard population with a fixed age distribution - such as the World standard by Segi, later modified by Doll - is applied to the age-specific rates to obtain an expected summary rate if the population of interest had the same age distribution as the standard (40). A new WHO standard world population has been proposed (112) but has been considered unnecessary for replacing the older standards (113). Example of the computation through direct standardization is provided in several sources, and where one can also inspect age-specific rates (40, 114) (also see Appendix 1).

In an ideal situation, age-standardized rates are accurate only in the absence of an interaction between age and calendar time (115). Trends in age-standardized rates are sometimes not the best way to examine cancer risk at either temporal or geographic level, because according to Gardner and Osmond (14, 15) , the age standardized rates by calendar period can mask important changes in the age-specific rates, particularly in the presence of strong generation (birth cohort) effects. According to these authors, there can be no substitute to an inspection of age-specific rates in temporal analyses, because these serve to validate the use of age standardization and provide background information in interpreting results from

more complex modelling procedures. However despite this critique, a lot of valuable information can still be teased out from visual descriptions of the age-standardized rates over time across populations worldwide (116).

2.5.1.2 Choice of scale: arithmetic or log-transformed

Devesa *et al*., while reviewing their own work on observed incidence trends of cancers (100) found that improvements can be made in graphical presentations by choosing either arithmetic or logarithmic scale in plots of rates over time (103). These two types of scales that are chosen according to research question to be addressed, one where no transformation is carried out on either x or y axis (arithmetic) and the other where logarithmic transformation is carried out on y coordinate. They noted that different presentations of the same data can result in different impressions, facilitating or hindering comparisons of trends in diseases over time, in particular when time-periods vary. In epidemiological literature, both transformations are regularly applied to graphically portray disease over time.

The use of arithmetic scales is appropriate where absolute changes in magnitude are of interest. It can be useful when temporal trends of greater magnitude are of greater importance and consequence (103). In public health programs, this could be, for example, the evaluation of costs for a vaccination programs over time, or changes in childhood vaccination usage, or predicted number of cancer cases that can be used as an indication of the resources required for treatment.

There are reasons why the logarithmic transformation may be of greater utility in studies investigating changes in rates with time (102, 104). First, rates of very different order of magnitude may be plotted and visually interpreted. A visual interpretation of rates associated with few events alongside those with many events is enabled, such as trends in five-year age-

specific rates of lung cancer incidence in women aged 30-74, or the comparison of trends in rates of a rare cancer versus a common one. Secondly, it should be considered important to identify proportional changes in rates among populations where the baseline rates differ (103). The log transformation is particularly effective at depicting relative changes in risk over time in low risk relative to higher risk groups, providing evidence of similarities or differences in the trends between groups (105). Figure 2-1 illustrates these points by comparing the arithmetic and log scales in depicting age-standardized trends in female lung cancer incidence in five populations of US, Denmark, Australia, Japan and Spain. The uniform increases in Spain (No. 5 in Figure 2-1) compared to Japan (No. 4 in Figure 2-1) are played down by the use of arithmetic scaling of the Y axis (left side of Figure 2-1). On a log scale (right side of Figure 2-1), the increase in incidence rates in Spain is apparent i.e., a steeper increase in Spanish women is observed compared to Japanese women.

Both arithmetic and logarithmic scales have advantages in certain circumstances of interpretations; it has been critiqued, for example, that use of the log scale can conceal important generation effects, primarily in the older age groups when the rates are of a large magnitude (107). Sufficiently detailed labelling of a log-transformed axis may also be important, so that the absolute changes over time can also be evaluated. The log-scaled ordinate may be labelled 0.1 to 1, 1 to 10, 1 to 100, and so forth. If the rate of change is constant, the observed time trend will be a straight line on a log graph.

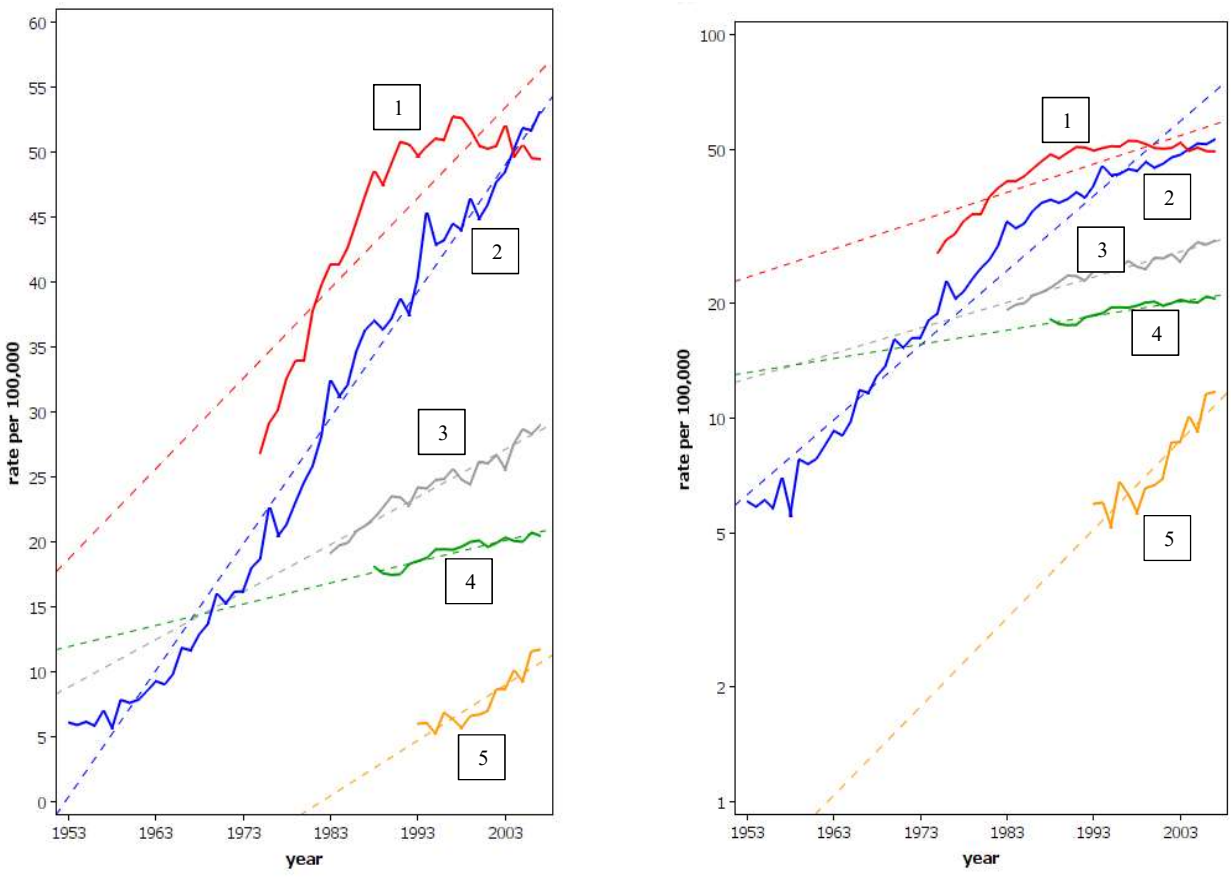


Figure 2-1: Age standardized incidence rates of lung cancer in women (15-85+ years) in five countries from 1953-2003 on arithmetic on left and log scale on right (source WHO- IARC).

1. United States
2. Denmark
3. Australia
4. Japan
5. Spain

Dashed lines represent general trends

2.5.2 Role of statistical models to quantify temporal change

It is sometime difficult to make conclusions based on graphical depictions only, because they are sometimes not straightforward and may not provide satisfactory levels of inference by themselves. Therefore, in these kinds of situations, our understanding of the progression of cancer risk can be greatly improved using more formal statistical procedures. Models offer quantitative and comparable estimates of trend based on objective criteria for choosing the best description of the data, and statistical tests to decide whether the trends may be real or due simply to chance (11). The consequences of subjective conclusions based exclusively on graphical descriptions can therefore be avoided. Although, the interpretation of cancer trends is however often complex and statistical models will not always provide definitive answers, but when used skillfully, they may however help in the interpretation of the observed temporal patterns.

Time trend data should be analyzed according to the problem under investigation, and the structural characteristics of the data. In cancer surveillance and monitoring, the objective of analysis might be to quantify recent trends in cancer, and to make statements as to the needs for future health priorities on the basis of anticipated future trends. In such a case, a diffident form of analysis e.g., *analysis of annual percentage change* provides a summary of the magnitude and direction of the trend, and can be obtained from *linear regression modelling*. The estimated annual percent change through linear modelling has an arbitrary element however, in that the trend estimate will depend on the extent of linearity in the selected period. Alternatively, one may wish to determine the annual percentage change for periods of time between abrupt changes in the linear trend e.g., when there are elements of *curvature* in the trend, and in which case, the estimated annual percent change will give imprecise

estimates of the average unit change. If and when this scenario happens, the alternate could involve modelling these abrupt or sudden changes in the trends, and the estimate the direction and magnitude of change for each period of time where the rates are relatively stable. Modelling methods to determine abrupt changes in the trend (slope) have been previously devised using weighted piece wise linear regression (117) and have been implemented using joinpoint method (118, 119). The idea behind the joinpoint regression modelling is that the linear trends should be derived over few continuous linear phases. The merits of this technique was demonstrated by Bray et al in their analysis of endometrial cancer incidence and mortality (7).

2.5.3 Other modeling techniques

To provide the researchers with clues to the etiology of disease, time trends are often jointly considered using age, time of event and date of birth. This approach known as age-period-cohort analysis, has been used in several studies by Bray *et al* (120-122), and involves fitting of age, period and cohort as explanatory variables in a linear Poisson regression model of the number of disease events, and offset by the corresponding person-years. This type of modelling, although complex, has been used for the temporal analysis of disease rates. The idea behind this modelling technique stems from the concept of "cohort-effect" that is also referred as "generation effects" (123). The cohort-effects are conceptualized as variation in the risk of a health outcome according to the year of birth, coinciding with shifts in the population exposure to risk factors over time (124). Cohort analysis began as a descriptive tool to better understand mortality, mostly for the purpose of forecasting and calculating life expectancy (125). For a long time since the original publication in Lancet by Kermack *et al*., in 1934 (126) , the definition, identification, and interpretation of cohort effects have been a

subject of controversy (127). To define a cohort effect, it is necessary to first define the related effects associated with age and period. Age effects describe the common developmental processes that are associated with particular ages or stages in the life course. In other words, age effects represent accumulated exposure, and/or the physiological changes associated with the process of aging. Period effects are the result of widespread environmental changes, the ubiquitous, population-wide exposures that occur at a circumscribed point in time (124).

According to Keyes *et al*., age-period-cohort modeling strategies can be defined as statistical attempts to partition variance into the unique components attributable to age, period, and cohort effects (124). Regardless of conceptual definition, the majority of these modeling strategies developed over the past thirty years assume that cohort effects can exist independently of age and period effects. Therefore, age, period, and cohort are often modeled as having a linear relationship with the outcome of interest, and each linear slope is estimated controlling for the effect of the other two. However, no statistical model can simultaneously estimate age, period, and cohort effects because of the collinearity among the three variables. This collinearity results in a statistically non-identifiable design matrix, making simultaneous modeling of the linear functions of three effects impossible.

Studies of international variation in cancer incidence are not possible without a large body of data to facilitate the detailed analyses of geographic patterns of cancers and their causes. In the next chapter, we present a brief overview of the international cancer registries that are the main sources of data on incidence rates, and the process of registration, and limitations in interpreting registry data. This is followed by Chapter 4 which discusses the main issues of cancer case ascertainment, and in particular completeness in a cancer registry.

Chapter 3 Background review on cancer registration

3.1 Historical context of cancer registration

Tabulating the number of people with disease is at the core of public health surveillance and descriptive epidemiology. People with disease are identified with reference to specific geographic regions and specific years; data on their lifestyle characteristics are often collected as well as their familial antecedents. The concept of counting health-related events evolved over time, and this progression can be followed in several key papers that underscore historic developments in surveillance and cancer epidemiology (128-134).

3.1.1 Disease surveillance: recording of complete and accurate data

According to Hippocrates, the endemicity of diseases and epidemic conditions are determined by the nature of a certain place, while climatic and behavioral elements are seen as the key forces driving them (129, 135). This supports the concept of collecting data on place, the natural environment, and people living in them for determination of the incidence of illness. Some 600 hundred years ago, with the emergence of scientific thought during the Renaissance, the concept of recording data on mortality and morbidity as a basis for preventive actions arose in Europe, and subsequently spread to the British Isles and Americas (131). The fourteenth and fifteenth centuries saw a primitive form of surveillance in Europe, which led to the first public health preventive measure by the Venetian Republic, in the form of detaining travelers from plague-infested areas (136). In the first half of sixteenth century, prompted by geographic ravages of the plague, records of vital events were maintained in European towns, and the "London Bills of Mortality" were published under the tutelage of the English Crown by elderly epidemic-scarred women known as "Ancient Matrons". How the information was used at the time is not recorded and it was not until one hundred years

later around 1665, when John Graunt, a London businessman subjected decades of data from the Bills of Mortality to detailed mathematical analysis. He was the first to estimate the population of London and to count the number of people who died from specific causes. He was also the first to conceptualize and quantify the patterns of disease, and to understand that numerical data on a population could be used to study the cause of disease (131, 132, 137). The need for more complete and accurate mortality data led to the establishment of the General Registrar Office in the United Kingdom in 1836, where the modern surveillance system originated under William Farr, a physician who was appointed as its first "Compiler of Abstract (medical statistician)". Farr was acutely aware of the necessity of matching registered "numerator data" with appropriate census "denominator data". He also influenced the authorities of the time to include a complete enumeration of the ages of people in the population census. The importance of carefully analyzing and continually reporting results of analyses on mortality rates enabled Farr to attract public attention to disease surveillance (133, 134, 138).

3.1.2 Cancer surveillance: inception of cancer registries

The registration of cancer patients is an important form of surveillance that cuts across different types of cancers. Cancer statistics were first compiled on a large scale in London in 1728. The practice of registering specific types of cancers for people employed in certain industries originated in England and Wales. For example, epithilomas due to tar pitch, bitumen, mineral oil, paraffin and soot were notifiable under the factory act of 1895 (139). According to Wagner, the historical development of the population-based process of registration was slow, and spanned several hundred years with many detours and blind alleys (75). It is however clear, that the earliest cancer registries covering defined populations used

multiple-source reporting. They were established first in Hamburg in 1927, and then in New York (1940), Connecticut (1941), and Denmark (1942). These registries are considered the oldest examples of modern cancer registries, and used information from multiple sources such as hospital patient records, diagnostic reports from pathology, and death certificates (27). Today, most of the developed and longstanding cancer registries have been linked with several relevant computerized databases, thereby enabling the capture of information of cancer patients beyond the traditional goal of registry dataset such as health insurance records, hospital information systems, and the census. In Section 3.1.3, we describe the historical role of specialized cancer organizations in the world-wide development of cancer registration.

3.1.3 Role of cancer organizations in cancer registration

3.1.3 (a) International Agency for Research on Cancer (IARC)

Although the first few registries in the 1940s were valuable, they did not cover much of the world population and they operated with few resources. An early impetus for establishing cancer registries came from a conference in Copenhagen in 1946 convened by the Director of the Danish Cancer Registry, Johannes Clemmesen. Twelve leading experts in the field of cancer control who attended this meeting made four salient recommendations: 1) cancer patients' data should be collected from as many different countries as possible; 2) the recording of data must be done in a comparable manner; 3) a central cancer registry in each country should be established to carry out local data collection and registration; and finally 4) an international body should be set up the responsibility of which will be to analyze the data obtained in each country. A World Health Organization (WHO) subcommittee on the registration of cancer cases was established in 1950 and provided the first set of

methodological guidelines for cancer registration (75, 139). The same year, at the International Symposium on Geographical Pathology and Demography of Cancer, organized by the Union for International Cancer Control (UICC) in 1950, the need for enumerating all new cancer cases in a defined geographic area was emphasized (140). From this landmark initiative, the first volume of the Cancer Incidence in Five Continents (CI5-Vol I) emerged as a technical report of UICC, the publication of which gave an immense boost to cancer registration activities at an international level (141). In 1965 (75), the Interim Commission of WHO recommended the establishment of national cancer registries (142, 143). In 1965, the International Agency for Research on Cancer (IARC) was established in Lyon, France, as a specialized cancer research center of the WHO. It began its work in July 1966. The governments of France, the Federal Republic of Germany, Italy, the United Kingdom and the United States agreed to become founding participating countries, each contributing US$ 150,000 a year towards IARC. Very shortly afterwards, the Soviet Union joined, followed by Australia and the Netherlands. Between 1970 and 1982, Belgium, Japan, Sweden and Canada added their support to the agency, and in 20 years' time, in 1985, the budget rose close to US$ 9 million.

3.1.3 (b) International Association of Cancer Registries (IACR)

Because of the creation of IARC in 1966, following a meeting in Tokyo in the same year, the International Association of Cancer Registries (IACR) was founded. The IACR functions as a membership forum for cancer registries of different countries, and is involved in the development of standards for cancer registration, training of registry personnel, publication of registry data, and holding of scientific meetings. After successful publications of the first two volumes of CI5 in 1966 and 1970 (26, 144), the UICC requested IARC to take full

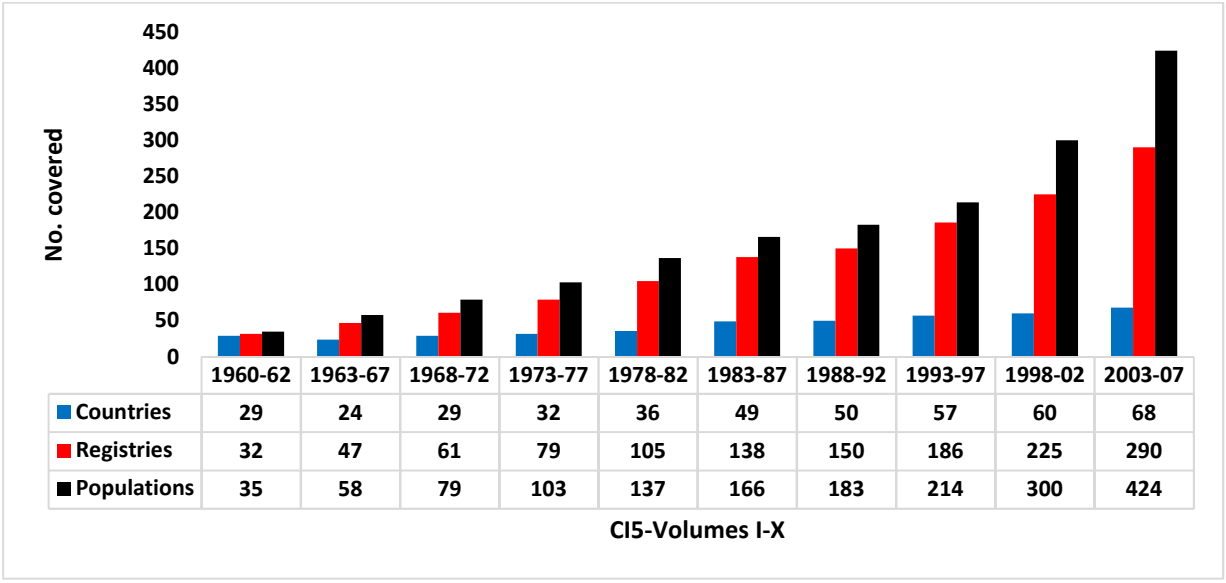
responsibility for its future productions. The alliance between IACR and IARC was therefore natural and resulted in publications of eight more volumes of CI5, the most recent volume (CI5-vol X) being published in 2013.

3.2 Cancer Incidence in Five Continents (CI-5 & CI-5Plus)

The overall objective of this resource is to provide *comparable* data on cancer incidence from different geographic regions of the world over time, while adhering to various criteria of quality that are gradually evolving in complexity (145). Each volume of CI-5 represents an interval of approximately 5 years of time (Figure 3-1). An additional dataset known as CI-5plus provides annual incidence rates for the longest time period in addition to 5-year interval of CI-5. Cancer registries from Denmark, Norway, and Sweden are the longest surviving registries reporting incidence rates annually since 1953.

Figure 3-1: Coverage of cancer registries in ten volumes of CI-5

Source: Bray et al, 2015 (40). Population signifies subpopulations within a single registry e.g., ethnicity, urban-rural etc.



	1960-62	1963-67	1968-72	1973-77	1978-82	1983-87	1988-92	1993-97	1998-02	2003-07
Countries	29	24	29	32	36	49	50	57	60	68
Registries	32	47	61	79	105	138	150	186	225	290
Populations	35	58	79	103	137	166	183	214	300	424

While the aim of the CI-5 is to report cancer incidence from as many regions of the world as possible, not all of them are included in the final volume (Table 3-1). The editorial process includes a careful check of submitted datasets from individual registries to ensure that the minimum criteria of quality are followed. For respective cancer registries, publication of their dataset in CI5 is seen as recognition of their advancement and quality, and many use this achievement to secure funding for their registration activities (27, 40, 145).

Table 3-1: Numbers of registries submitted and number and proportion finally included in Cancer Incidence in Five Continents Volume X, by continent.

Source: Bray et al, 2015 (40).

	Registries submitted	Registries included	Proportion accepted
Africa	18	8	44
Asia	102	63	62
Central & South America	35	25	71
Europe	136	118	87
Oceana	11	10	91
North America	70	66	94
Total	372	290	78

3.3 Hospital-based versus Population-based cancer registries

Cancer registration can either be hospital-based or population-based (22, 146). Although the process of registering cancer cases is identical to a large extent in both types of cancer registries, a clear distinction must be made.

3.3.1 Hospital-based (and Pathology-based)

These registries record cancer cases in hospitals, usually without the knowledge of background population (i.e., these registries cannot provide measures of the occurrence of cancer in a defined population because it is not possible to define their catchment populations, or the populations from which all the cases arise). Hospital-based cancer registries are concerned with the recording of information on the cancer patients seen in a particular hospital (147). The emphasis is on clinical care and hospital administration. Hospital-based registries serve most often as a central source of data representing the population in which the hospital is located (146, 148). This type of registry can calculate the frequency of cancer cases and measure the outcomes for the patients it treats. It can provide useful information that can guide resource allocation, cancer care policies, and investment in cancer prevention activities. However, a hospital-based registry cannot provide incidence rates because the denominator "population at risk" is not known (149). It can still be useful in describing referral patterns and can therefore be helpful in defining the catchment area of a given hospital (150). Another similar type of registry is pathology-based, which records cancer cases diagnosed in pathology laboratories based on histopathology and cytology reports.

3.3.2 Population-based

Population-based cancer registries are the gold standard for calculating true or close to true cancer incidence in any given population. They are very resource intensive compared to hospital-based and often require long-term state- or region-specific support to operate. This type of registry collects information on cancer patients from various sources, e.g., hospitals, pathological laboratories and death certificates, and provides estimates of the magnitude of

cancer occurrence in a specific geographical area. Methods of data collection (*notification methods*) have traditionally been classified as *active* or *passive*. Active reporting involves registry personnel visiting various sources, abstracting the required information onto special forms and obtaining copies of the necessary documentation. Passive (or self-) reporting relies on other health workers to complete notification forms and forward them to the registry. In practice, a mixture of these two systems may be used, with, for example, active hospital visits being supplemented by passive receipt of copies of pathology reporting-forms and death-certificates mentioning cancer (151, 152).

Chapter 4 Cancer case-ascertainment in cancer registries

Cancer registries vary according to population size, funding, and trained manpower available for functioning. Most of these registries have strategic and logistical autonomy and follow their standard procedures accordingly. There is therefore a possibility that this autonomy affects the quality of registered data. The usefulness of population-based cancer registries across different geographic regions depends heavily on quality indices of registration and in particular, on *completeness* (150, 153-157). Cancer incidence rates can be misleading when a cancer registry cannot meet standard procedures for data abstraction and coding. The issues of cancer case-ascertainment and its completeness are followed up in this chapter to support the rationale of carrying out the current thesis work.

4.1 Overview of cancer case-ascertainment

Cancer case-ascertainment is commonly called *case-finding*, and is the *process of identifying patients with malignant cancer who meet the inclusion criteria for a cancer registry*. Because cancer registration requires continuous monitoring of cancer incidence and mortality, case-ascertainment must be carried out by identifying *all* cases in a defined population, regardless of where the cancer patient was identified in the healthcare system. Common sources of case-ascertainment include hospitals, independent treatment centers, pathological-laboratories, physician offices, coroner's offices, health insurance systems and nursing homes (149, 152). In hospitals, medical records are coded and indexed, making them the principal source of case-ascertainment. These codes permit the retrieval of records related to reportable cancers that must be included in the registry. Reportable cancers are the ones that meet the inclusion criteria of a registry. These cancers are well-defined in the International Classification of Disease for Oncology (ICD-O), a coded nomenclature

published by the WHO. The ICD-O defines each type of tumor and its behavior, as benign, uncertain malignant potential, in situ, invasive, or metastatic. The reportable cancers collected by all cancer registries are those that are malignant (in situ and invasive). Metastatic tumors (malignancy growing on a site at a distance from the origin in which it started) are not reported individually; rather, metastases are reported as a progression of the tumor at the site of origin. Occasionally a registry will require that certain types of benign tumors be reported, such as benign brain tumors, which cannot spread but do have the potential to be lethal, and tumors of uncertain malignancy, such as carcinoids of the appendix (158).

4.2 Key issues in cancer case-ascertainment

Goldberg (159) and Donaldson (160) reviewed evaluation methods for registries and listed some key issues to be addressed before setting up any kind of registry. For a cancer registry, this means addressing fundamental issues of ascertainment such as comparability, validity (or accuracy), and timeliness (16, 17). In this section, the quality indicators regarding cancer registration are briefly addressed.

Comparability is the extent to which "classification" and "coding" of new cancer cases, together with the "definitions of incidence" adhere to international standards and guidelines. The comparability of statistics produced for different registries across various regions, and over time, is crucial to a meaningful interpretation of incidence trends. The standard for classification and coding of cancer is ICD-O (16, 40, 151).

Validity (or accuracy) is defined as the proportion of cases registered with a given characteristic (e.g., sex, age or diagnosis) which truly have this attribute. It depends on the veracity of source documents and the level of ability in abstracting, coding and recording this

information in the registry database. Bray and Parkin (16) have described four methods to evaluate the validity of cancer registry data: the diagnostic criteria method i.e., *histological verification* (percent of cases with morphologically verified diagnosis: MV%;) and *death certificate only* (percent of cases for which the only information came from death certificate: DCO%); missing information analysis; re-abstracting and recoding; and the internal consistency method. These methods provide numerical indices of validity as well as for completeness.

Timeliness in the reporting to the cancer registry relates to the rapidity with which a registry collects, processes and provides sufficiently reliable cancer data (16, 161).

Completeness is the most important attribute and the key quality indicator of any cancer registry. It is defined as the *extent, degree or proportion of all incident cancer cases in a defined population that is included in the cancer registry database*. In theory, all cases of cancer in a defined population should be recorded in a population-based cancer registry or should be as close to 100% as possible.

Different methods are sometimes employed to measure and evaluate completeness in different cancer registries, which means that some registries could differ in quality in terms of completeness in case-ascertainment. In the next section, some of the available methods of completeness of cancer case- ascertainment (also known as completeness indices) in cancer registries are discussed.

4.3 Disparate methods of evaluating completeness

In the past, several methods were used to assess completeness of case ascertainment in cancer registration (153, 154, 159, 162-167). Parkin and Bray have separated these methods

into two broad categories (17). "*Qualitative (or semi-quantitative) methods*" give an indication of the degree of completeness relative to other registries, over time e.g., Historic data methods, Percent of morphologically verified cases (MV %), and Mortality-to-Incidence (M:I) ratios. "*Quantitative methods*" (such as death certificate methods e.g., percent identified by death certificate only (DCO %), capture-recapture methods, and the Bullard Flow method) provide a numerical evaluation of the extent to which all eligible cases have been registered. Some of these methods are briefly described:

Histological verification of diagnosis:

Histological verification of cancer or *"Percent of cases morphologically verified (MV %)"*, is a measure of the validity of the information and completeness in a registry (25). A very high proportion of cases diagnosed microscopically by histology or cytology/hematology (higher than reasonably expected) suggests over-reliance on the pathological laboratories as a source of information, and failure to find cancer cases diagnosed by other means. The percentage of cancer cases likely to be histologically verified for a given site is dependent upon local regional circumstances where the registries are situated (40). It might be low if the means for taking biopsies, or examining the tissue, are lacking or inadequate e.g., in low resource countries. Conversely, the availability of sophisticated imaging techniques may reduce the need for biopsy.

Mortality:Incidence (M:I) ratios:

The M:I ratio is a key indicator of completeness and involves comparison of the number of deaths (obtained from a source independent of the registry, e.g., the vital statistics system) and the number of incident cancer cases, registered in the same period (25). The M:I ratio

may also reflect local conditions because survival and the quality of mortality statistics are at many levels related to the socioeconomic development of the region. Values of M:I greater than expected, i.e., exceeding 1, signals under-registration (incident cancers missed by the registry), and becomes more noticeable if it involves more than one type of cancer in a registry. However, under- or over-reporting of tumors on death certificates distort this ratio, as will a lack of constancy in incidence and case fatality (the rate of death amongst incident cases) over time. Application of this indicator of ascertainment does require, however, mortality data of good quality especially concerning accurate recording of cause of death (168).

Death certificate methods (DC methods):

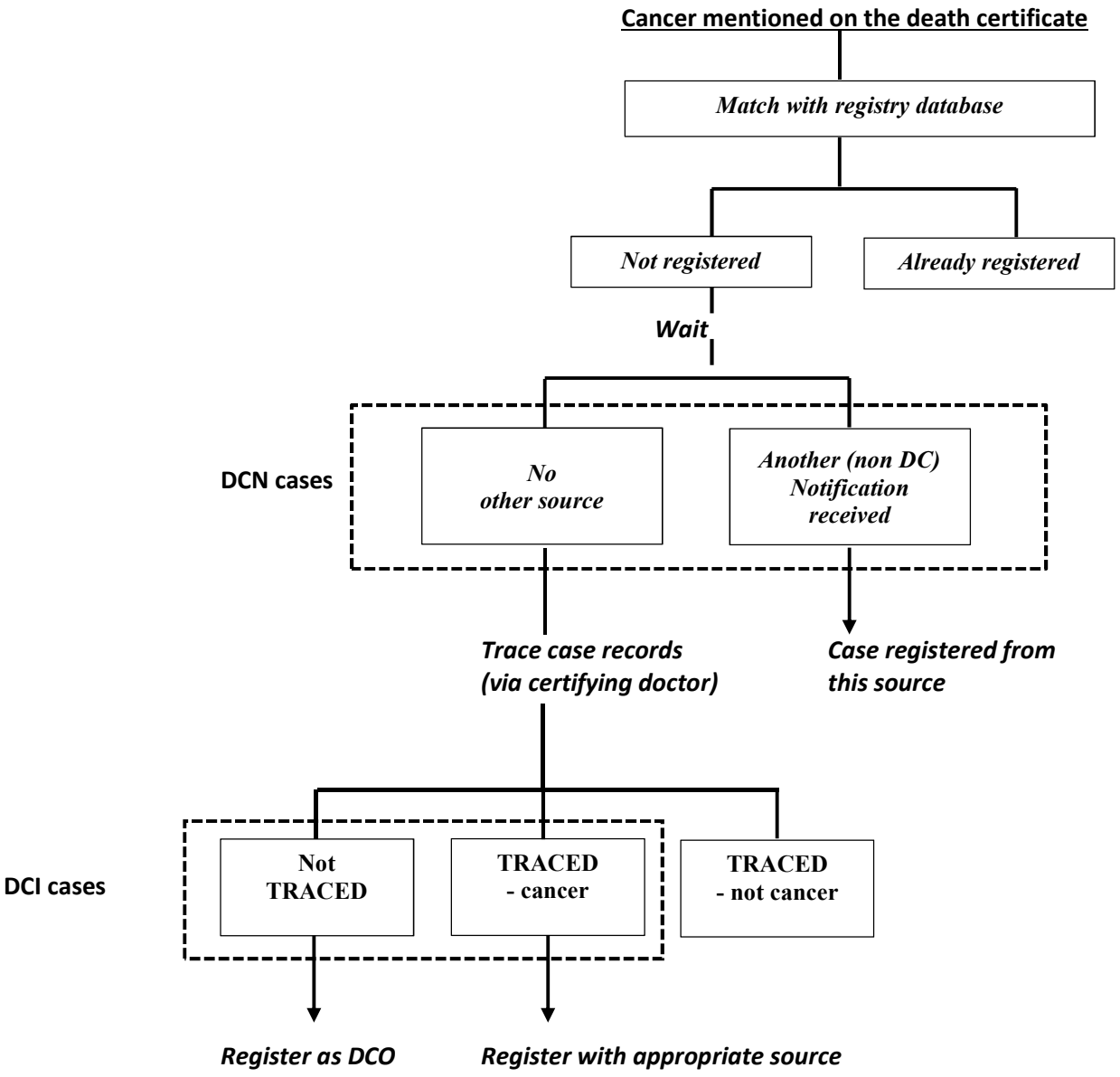
Death certificates are one of the main sources of information in a cancer registry in developed countries (169), and have three main uses in cancer registration: 1) as a complementary source of information on new cancer cases, 2) as quality control assessment of both completeness and validity, and 3) for studies on survival of registered patients. DC methods cannot be readily applied to cancer registries from low- and medium-income countries (28).

Methods used by Ajiki (170) [and quoted by Parkin (17) and Kamo (171)] explains death certificates as a means of capturing information on cases that were not registered during life. A "*Death Certificate Initiated*" (DCI) registration is one for which any information available from other sources was found as a result of trace-back procedures, initiated because of a first information via death certificate (Figure 4-1). It is important to note that DCI cases are different from DCN (*Death Certificate Notification*) cases, the ones for which subsequent information was received without the need of a trace-back (follow-back) enquiry (Figure 4-

1). After all the trace-back procedures performed on DCN cases, the cases for which no other information than a death certificate mentioning cancer was obtained are termed DCO (*Death Certificate Only*) cases. Therefore, although the DCO case is not an ideal indicator of completeness of registration, an elevated proportion of cases diagnosed through DCO (%) is suggestive of incompleteness.

Figure 4-1: The use of death certificates to identify new cases of cancer.

Source: Parkin and Bray, 2009 (17).



Bullard's Flow method:

Flow method, was introduced by Bullard in 2000 (153) and is based on the logical flow of the data in the registration system, and on the time distribution of various probabilities inherent to this flow, which can be calculated using routine cancer registry database. To estimate completeness of cancer registry data, the number of patients not registered at a given time after diagnosis of cancer has to be ascertained. These patients are divided into two groups: (1) patients who are alive and still unregistered (missing) and (2) patients who have died without being registered during life, and remain unregistered because cancer was not mentioned on the death certificate (lost) (172).

Capture-recapture methods:

Capture-recapture methods were originally conceived as a tool by ecologists to estimate the size of free-living animal populations in a close environment. In this method, samples of animals are captured, tagged, released and then recaptured, and the size of the animal population is estimated from the numbers of animals captured and recaptured in each sample (173). Since cancer registries use multiple data sources, capture-recapture methods can be applied to these incomplete registry of patients to evaluate completeness of registration. Such sources may include notifications by clinicians, radiologists, pathologists or through death certificates (174). The simplest capture-recapture method involves two sources of information as shown in Table 4-1. This method can also employ more than two sources to estimate completeness of a case-ascertainment in a registry (17).

Table 4-1: Capture-recapture method using two sources of information of cancer case ascertainment.

Source		**Source A (e.g., pathology reports)**	
		Yes	***No***
Source B	***Yes***	n_{11}	n_{01}
(e.g., death certificate)	***No***	n_{10}	n_{00}

n: cases identified through 2 sources; pathological considered as 'source A', and death certificate as 'source B'.

There are two assumptions behind the capture-recapture analysis of completeness method (175): (1) There is no dependency in the capture process between all sources in a multi-source model e.g., when there are two sources of identification such as in Figure 2-2, identification or capture of cancer case by one source A is independent of capture by second source B; and (2) All individuals with cancers have the same probability of being captured.

Parkin and Bray have provided examples of violations of these assumptions in their review on completeness of cancer case ascertainment (17). Cancer cases captured by one source may have higher, or lower, probability of being captured by the others leading to dependence (violating the assumption no. 1): terminal patients are less likely to be admitted to the hospital and therefore not appear on hospital discharge report (source A), and more likely to die and have a cause of death labelled as cancer on death certificate (source B), thereby creating dependence between two sources. It is also possible that some individual characteristics are related to the probability of capture (violating the assumption no. 2): patients living near the border of registration area may go to hospitals outside of it, and therefore be missed by hospital resources.

Silcocks and Robinson (176) have pointed out issue of sample size in capture-recapture method: that while small sample sizes result in wide confidence limits, large sample sizes are

inoperable because of the workload concerned. Several cancer registries have applied the capture-recapture method successfully (174, 177) and studies have encouraged a careful application of this method to supplement the traditional ones (178, 179). There are ways available to deal with both the problem of dependency between sources and the problem of the characteristics associated with capture (163, 167, 177).

Further details on advantages and disadvantages of these different methods in different cancer registries are available (28, 168, 171, 174, 176-181).

4.3.1 Comparison of completeness indices in literature

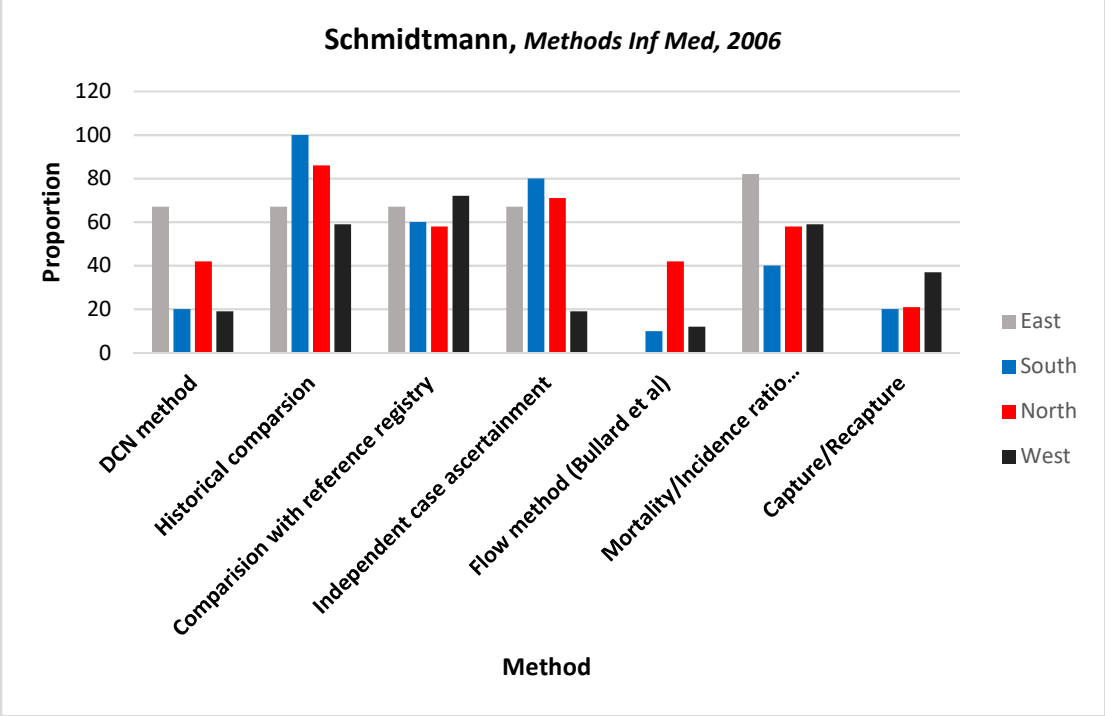
Both standard indices (MV, DCO, and M:I) as well as complex sophisticated methods (Bullard's Flow and capture-recapture method) are used in estimating the degree of completeness of ascertainment. This means that with the availability of various methods, the degree of completeness in cancer registries will vary with whatever methods are used. Schmidtmann and Blettner carried out the first survey of its kind to compare different methods that European cancer registries use to assess completeness of ascertainment (166). The study revealed that 86% of the 56 cancer registries that returned the survey questionnaire (of total of 195 registries that were contacted) had complete case ascertainment. The methods used most frequently were comparing current with historical incidence (73%) and comparisons with a presumably complete reference registry (65%). The mortality-to-incidence (M:I) ratio was used in 58% of registries. More complex procedures, such as the capture-recapture method (25%) and Bullard's flow method (21%), were employed less often. The use of more than one method was also infrequent (29%). Zanetti et al. repeated the survey in 2015, with an improved response rate of 65% from cancer registries in Europe (161). The methods used were still largely based on simple indices with only slight

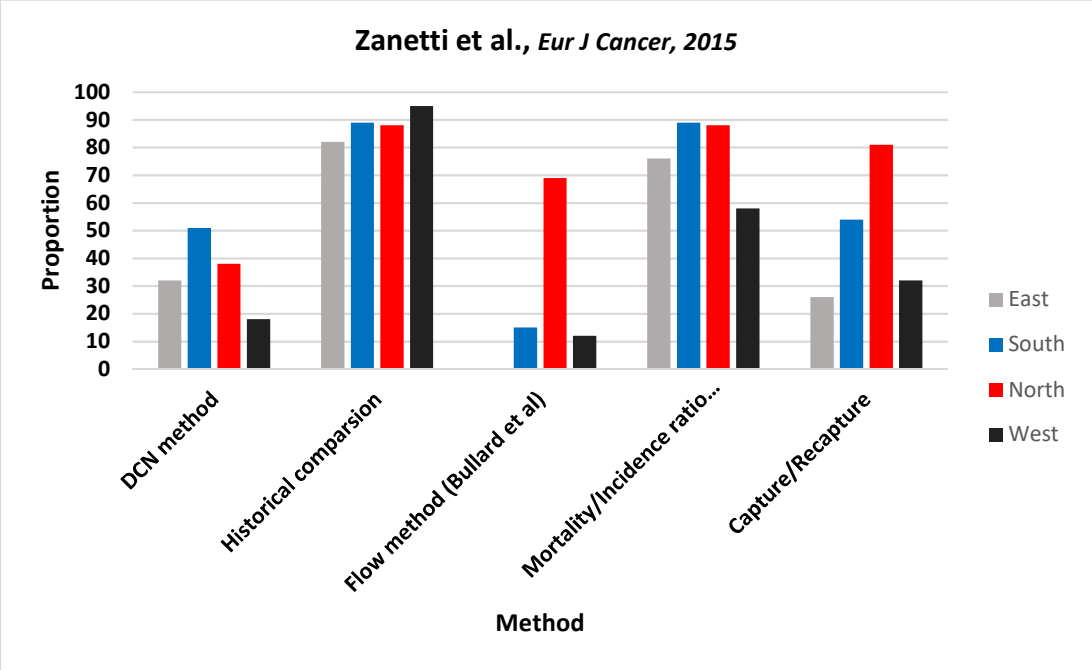
improvement in the use of quantitative methods. The proportion of cancer registries within each region of Europe, using various methods for estimating completeness from surveys carried out by Schmidtmann and Zanetti respectively, are shown in Figure 4-2. The impression gained from these surveys is that there are very different methods in use by individual cancer registries, and there are few comparative studies on their performance in relation to ascertainment in males and females. The authors of these studies (161, 165, 166) suggest that in order to make valid comparisons across regions, modern registries should work more on standardizing methods of estimating completeness.

Most studies on methods of estimating completeness of ascertainment which address the sources of imperfect ascertainment (16, 17, 161, 166) have detected a "country-effect" in missed cases that became recognizable with use of more sophisticated methods. Cancer registries in countries that are well funded and cover large segments of the population are likely to use more sophisticated methods like capture-recapture and Bullard's Flow method. On the other hand, cancer registries that are restricted to smaller populations with limited funding use simple methods or standard indices of data quality and completeness e.g., MV, DCO, or M:I ratio, among others. In addition to assessing different methods used in different registries, there are few studies comparing performance of these methods in males and females. Castro *et al* attempted to assess the merits of more advanced methods for completeness of ascertainment. However their study was limited to relatively smaller regions in Europe and assessed only one type of cancer (169, 173).

These studies, however, have underscored the importance of unifying methods for estimating completeness that could improve validity of incidence comparisons between cancer registries in both males and females.

Figure 4-2: Surveys results on proportion of cancer registries within each region of Europe using different methods of estimating completeness. (a) Schmidtmann and Blettner, 2006 (b) Zanetti et al in 2015.





Countries grouped according to the definition of the UN Population Division (East, South, North, & West).
DCN method is from where the no. of cases come from death certificates only (another term for DCO %)

4.4 Completeness of cancer case-ascertainment by sex: Is there a gender-bias?

Global collation of data on new cases of cancers through population-based cancer registries provide an excellent opportunity to explore sex differences in cancer incidences across diverse geographical regions (182). These differences are interpreted in light of genetic and environmental causes of cancers *within* and *between* countries (39, 183).

Much has been written on sex disparities in specific types of cancer in both developed and less-resourceful parts of the world (184-190), yet there is severe scarcity of literature on the possible under-registration in cancer registries according to sex, in particular among women. Since the decade of 1990s, there had been an increasing call to systematically quantify the level of completeness of cancer registries in the region in which they operate (31, 162, 191). That call was heeded in the decade of 2000s, when studies on completeness of registration from Africa and Eastern Europe started appearing in literature (156, 192) in addition to developed parts of the world (180, 193). There are only two studies that discussed and showed possibility of under-reporting in females in a cancer registry (182, 194). Barlow *et al*. found overall under-reporting of 3.7% in a Swedish Cancer Registry for the year of 1998 (194). In their study, there seemed to be a pattern of under-reporting that was worse in women among elderly, and among men in younger age group. Pearce *et al*. (182) concluded that the underlying socio-economic level of the society and its culture in developing countries should be borne in mind when interpreting incidence rates of childhood cancers because girls may be more likely to be under-diagnosed than boys. In some cancer registries, the extent of miscoding and misclassification by sex has also been noted, albeit to minimum (42).

Therefore, in light of studies mentioned above, it is reasonable to suspect that in some resource-poor countries and conservative societies, due to socio-cultural dynamics, a female cancer patient may be omitted from the cancer register. This will have important implications in the incidence statistics reported from the registry (183). It may be that females who are missed are different in some way in terms of incidence outcomes, from those who are identified by the registration process. Under this scenario, the incidence rates obtained from the incomplete data or through under-registration of cancer cases can be biased.

Quantitative assessment of this kind of gender-bias in registration was inferred using data from the Kampala Cancer Registry in Uganda by Templeton and Bianchi (195). Their publication in 1972 reported registration of females to be half as complete as those of males. However, they also reported that this bias in registration diminished as social patterns of literacy and health awareness evolved and hospitals became more accessible (195, 196). But even with better medical services and the progressive social emancipation of women over time, reports from cancer registries in developing countries are criticized because they represent a biased sample of prevalent cancers in the region. Moreover, universal healthcare remains a distant prospect in many low-resource countries, and even when improvements are seen on the horizon, coverage is not likely to be equal in both men and women (197). In addition to problems of health-care accessibility (more so in female patients), a cancer diagnosed in a hospital can also be influenced by age, tribal/ethnic affiliations, education, and social status in some countries (198).

Table 4-2 presents results of independent studies on cancer case-ascertainment in males and females using standard indices of completeness. The Canadian registry (199), by far shows best completeness indices relative to others in both males and females. These

population-based registries are included in CI-5 database with the exception of Gambian registry (200). Sweden is the most complete registry in all volumes of CI-5, with MV% of about 98 in both males and females in the most recent volume (25).

In Table 4-2, Gambian study clearly revealed heterogeneity in quality indicators, and the results strongly suggested incomplete ascertainment in both sexes. Lower incidence rates for several cancer types in both males and females were reported in comparison with other West African registries such as in Mali, Guinea, Cote d'Ivoire, Niger and Nigeria (200, 201). According to the authors, the differences between Gambians and other Africans in cancer incidence may either represent true geographic variation in risk or there might be other factors at play. One factor is the registry's predominant coverage of rural population of Gambia, and the related fact that other comparable registries in Africa are not rural. This rural-urban contrast highlights several possible issues (e.g., under-utilization of medical facilities in rural areas, under-diagnoses of cancer in low resource rural health care settings, and under-reporting of cancer cases from rural populations by registry staff), or it could really represent a true difference in the risk of cancer between rural and urban regions (in this case, a truly lower incidence in rural Gambia). Although the importance of each of these possibilities is difficult to judge, a similar urban–rural difference in cancer incidence in both sexes has been observed elsewhere (202, 203), and much of the difference was attributed to socio-economic deprivation. One study in the US described urban-rural gradients in cancer incidence in both sexes (204). Explanations for the type-specific incidence differentials in urban-rural gradients were more or less the same as those proposed for African registries (e.g., differences in exposure, cancer screening, access to physicians and facilities to obtain a timely cancer diagnosis, and differences between urban and rural areas in the characteristics

of the populations at risk including ethnic origin, population diversity, culture and health behaviors).

Table 4-2: Completeness of cancer case-ascertainment for all ages in males and females using standard methods of ascertainment.

PBCR	Authors (Ref #)	Male			Female		
		MV (%)	DCO (%)	M:I (%)	MV (%)	DCO (%)	M:I (%)
Bulgaria	Dimitrova, 2015 (205)	73.3	9.8	65.9	82.8	6.9	50.5
Canada	Zakaria, 2013 (199)	90.0	0.9	48.8	90.0	1.2	48.5
Spain	Navarro, 2010 (206)	88.7	2.6	52.3	87.8	3.8	48.0
Italy	Tumino, 2004 (207)	83.0	2.0	54.0	85.0	3.0	48.0
India	Mathew, 2011 (208)	83.2	1.4	12.6	81.5	1.1	9.3
Gambia	Shimakawa, 2013 (200)	18.1	6.6	NR	33.1	3.6	NR

MV%: percent morphologically verified; DCO%: percent death certificate only; M:I Mortality-to-incidence ratio

Completeness of cancer case-ascertainment, can therefore be confounded by gender-effects in terms of access to cancer care services. Access to health care services are considered basic human rights and yet these rights are not always distributed equitably among men and women in many parts of the world, not to mention absence of treatment services (209, 210). Evidence suggests that women with disabilities have worst access to preventive health services. Some of the studies on this topic are small-scale (211, 212), however they provide important insights into the experiences of women as they navigate the healthcare system. They do not, however allow any conclusions regarding utilization of preventive services at a population level. While the study of cancer care access by Sakellariou and Rotarou (209) focused on comparison among disabled and non-disabled women, their conclusion can be equally applied on the male-female differences in the access to health services e.g., poor socioeconomic conditions of women in their lack of utilization of cancer care services. A study by Tabatabai *et al.,* used longitudinal linear mixed-effects

model to demonstrate that age-adjusted incidence rates in lung cancer have decreased across US, but the gender disparities persist. Their model, however also predicted that the gender gap will gradually disappear by mid-2018 in Whites and by 2026 among Blacks. Gender effects were less clear in other studies, while some researchers have suggested that men receive more cancer detection tests than women in the same medical practices (213, 214). Lack of access to health care services in some parts of the world give some indication that there might be a gender gap in cancer registration as well. However, there are negligible number of studies (182, 215) that have embarked on exploring this issue in the field of cancer surveillance, specifically registration process itself. Parker (215) pointed out an exceptionally high cancer registration ratio (in boys relative to girls) for childhood cancers in Afghanistan, Bangladesh, Morocco, Pakistan, and Papua New Guinea; and concluded that striking gender-bias give more information on the socio-economic dynamics rather than the etiology of the cancer.

Regarding methods of ascertainment of cases according to sex, Castro *et al*. showed the applicability of more sophisticated methods in worst and best-case scenarios in terms of the availability of information from death certificates in the Portuguese Cancer Registry (169, 173). They chose gastric cancer for their study because it is, in both sexes, one of the most common and lethal cancers in the Northern region of country. The overall estimates for completeness of gastric cancer registration, using different methods, were approximately 80% rising to 95% in the best-case scenario. The results also showed that females had a similar completeness of registration, albeit somewhat higher (around 85%) than males (around 81%), for all methods (Capture-recapture and Flow method).

In summary, this chapter highlighted contextual problems that can exist when dealing with completeness issues in registration process in males and females regardless of whatever methods of completeness are used. These methods of completeness provide clues to the quality of cancer registries. Cancer cases can only be recorded once they have been diagnosed, after a patient has presented themselves to medical attention. It is possible that in rural areas of developing countries, people can die with their cancer never having been seen by a medical doctor. This might not be always true in the urban populations of the twenty-first century (27). In some countries, cancer registration has a legal basis and is funded by governments, but some registries, particularly in developing countries have operated on a voluntary basis, relying on good will and the tradition of sharing of medical information among different medical specialties (216). While acknowledging the existence of these contextual obstacles in cancer registration, particularly in terms of gender-bias, the population-based cancer registries included in CI-5 provide a good source of information to study the causes of cancers (25, 40, 41).

Chapter 5 Sex Ratio: a distinct approach in light of imperfect ascertainment

The preceding sections of thesis have highlighted 'completeness' as a major quality issue in cancer registries. Not only do different registries use different methods to assess completeness (from simple methods used by developing registries to sophisticated methods used by well-established registries), but a possibility of gender-bias could exist because of male-female inequality in access to health care services. Imperfect cancer case ascertainment due to use of different methods along with several artifacts as discussed in previous chapters, can lead to inaccurate estimates of incidence rates. Any geographic comparisons of incidence based on imperfect or incomplete ascertainment will be biased, and any inference based on variation in incidence patterns with time or across regions will not be convincing. We therefore suggest exploring the SR (i.e., ratio of male-to-female incidence rates) as a key parameter of cancer occurrence, as a distinct partial solution to imperfect registration, and as a distinct analytical approach in terms of it's variability, to conjecture on causes of cancers (assuming that imperfectness or incompleteness is relatively similar in males and females among peer approved cancer registries included in CI-5). In the following next sections, literature is reviewed regarding robustness of sex ratios methodology that has recently opened new perspectives in the epidemiology of diseases such as multiple sclerosis, Parkinson's disease, rheumatic diseases, infectious diseases, cardiovascular diseases, and cancers.

5.1 Literature review

5.1.1 Sex ratio: incidence of multiple sclerosis

Multiple sclerosis is an immune-mediated disease in which the body's immune system attacks the central nervous system. Studies from developed parts of the world, including the

USA, Australia, and Canada have shown the sex ratio (female relative to male) of multiple sclerosis cases to increase over time when consecutive cross-sectional comparisons were made (36, 217). Although the etiology of this disease remains mostly undetermined, yet both genetic and environmental factors have been hypothesized in several epidemiological studies (218, 219). There is a consensus in these studies that in many parts of the world, the incidence of multiple sclerosis is increasing, in particular, among women. According to Orton *et al*, this worldwide change in incidence is often attributed to inconsistencies in ascertainment of cases of multiple sclerosis (36). Using a population-based data from Canadian Collaborative Genetic Susceptibility to Multiple Sclerosis, they indicated that this increase was very likely to be real and not due to artifact related to changes in ascertainment. The speed of change in sex ratios of multiple sclerosis implicated the environmental causes and possible gene-environment interaction. Another study using the same database of Canadian patients indicated that one of the sub-type of multiple sclerosis (i.e., relapsing-remitting type) is perhaps more susceptible to environmental influences and might account for increase in number of female patients (220). These observations prompted the investigators who proposed sex-ratio methodology, to open up new line of investigations in the epidemiology of multiple sclerosis. Their rationale of using sex ratios was that the incidence of multiple sclerosis has varied in the literature and are partly related to the geography of the disease. According to Ramagopalan *et al* (220) and Orten *et al* (36), there seemed to be a general trend over the years towards increased incidence, however results in different countries were not easy to compare since some reported the increased incidence, whereas others showed no change or decreased incidence. Ascertainment of cases of multiple sclerosis also confounded multi-region comparisons in the literature. Their possible solution

of overcoming or reducing ascertainment bias included uniformly sampling the same population over time. Their justification was that by showing that changes have taken place in a subset of patients will enable them to use an internal contemporaneous matched control group, through the measurement of sex ratio.

In terms of hypotheses generation through the use of sex ratio (by birth year) of multiple sclerosis, a heightened risk for later born female children in large pedigrees with sibships often spanning two decades or more in birth timing was conjectured (36, 220). Year of birth came out as a significant predictor of sex ratio (female-male) of multiple sclerosis in binomial logistic regression model that was used in analysis. The studies recommended that sex ratio by year of birth could be used as a partial surrogate for incidence, since incidence of multiple sclerosis seems to be changing and is also difficult to measure. An important point that Orten *et al*., discussed in their paper was the increasing sex ratio in favor of females and the issue of gender-bias (i.e., a relative underrepresentation of women) in countries where the natural sex-ratio (i.e., of the population and not disease) was also intriguingly low. This, according to them could serve as an important clue to the pathogenesis of sex difference in risk, and to the nature of environmental risk factor in the multiple sclerosis.

Several other studies have explored the environmental effect causing the changes in sex ratio of multiple sclerosis and have attributed these changes in lifestyle factors of women. These include higher numbers and changing roles of women in the workforce, outdoor activity, dietary habits, and alterations in menarche and in the timing of childbearing years, among others (218, 221, 222). In their, in-depth analysis of Canadian and Danish registries of multiple sclerosis and corresponding national statistics on smoking prevalence, Palacios *et al*., found that female-to-male sex ratio in smoking prevalence increased in parallel and was

strongly correlated with sex ratio of incidence of multiple sclerosis in both Canada and Denmark. The stronger correlation of intra-country analysis as compared to inter-country analysis suggested that factors other than smoking contributed to international variations in sex ratio of incidence of multiple sclerosis (32). Changes in sex ratio using Oslo Multiple Sclerosis Registry were studied over a period of seven decades by Celius and Smestad (222), and they concluded that relative stability of male cases of multiple sclerosis over such a long period makes smoking, an unlikely cause of increased incidence in females. Hence divergent views exist in literature regarding hypothesis that smoking increases the risk of multiple sclerosis, and whether that could also explain the increasing incidence in females.

A meta-analyses of epidemiological data on multiple sclerosis highlighted increasing incidence trends over the past few decades across different parts of the world (223). Koch-Henriksen and Sorensen (223) challenged the hypothesis of latitudinal effect in the disease incidence and suggested more focus on changes in lifestyle-behavioral patterns among women. The changes in sex ratio of incidence rates postulated that if a gene pool is stable, apart from adaptations caused by migration, any short term marked changes in incidence of multiple sclerosis can add weight to the notion of etiological environmental factors and will stimulate hypotheses formation. The investigators also discussed the notion whether the rising sex ratio of incidence rates (females relative to males) represents a true incidence rate and whether it could be due to better ascertainment of cases among women. The study considered incidence rate as a better measure of the risk of multiple sclerosis because it is not affected by insufficient ascertainment probability.

Trajano *et al*., however critiqued the results derived from the meta-analysis by Koch-Henriksen and Sorensen on the data collected at different times in distant regions and

countries (33). Theirs was a major methodological concern in regard to the variability in population size, composition, ethnic origin and age. Moreover, they stated that increasing frequency of multiple sclerosis could be ascribed, at least partially, in some countries, to a global improvement of case ascertainment and diagnosis being dependent on accessibility to adequate medical structures and personnel, and to the availability of more sensitive clinical tools which have been included in the new multiple sclerosis diagnostic criteria. They opined that because of this variability, it is very difficult to aggregate and compare data from different studies. However, despite their critique, they were supportive of using sex ratio as a new robust epidemiological approach to evaluate the changes in incidence and the geographical distribution of the disease. Based on International Multiple Sclerosis Registry, their results of changes in sex ratio over time demonstrated, unlike Kock-Henrikson and Sorensen, that the global increase of incidence is driven by latitudinal-effect i.e., sex ratio was stable in Southern latitude (Argentina, Australia, and New Zealand) whereas it was remarkable high in the Northern latitude (Belgium, Canada, Germany, Denmark, and Netherlands). A hypothesis regarding solar ultra-violet radiation exposure and its association with multiple sclerosis was proposed since this exposure varies with latitude and is also associated with biological activity of vitamin D. To date, this is the only large-scale study designed and conducted to directly compare the sex ratio trends over time among populations from different geographical areas. The latitudinal gradient hypothesis of multiple sclerosis was also confirmed by studies on sex ratios from regions in New Zealand (224) and France (225).

5.1.2 Sex ratio: incidence of Parkinson's disease

The studies on sex ratio of incidence in multiple sclerosis by Orton *et al*., in Canada and France (36, 225) was also an impetus for exploring sex ratio of incidence in Parkinson disease especially in terms of exploring the hypothesis that smoking reduces the risk of Parkinson's disease (226). Parkinson's disease is a neurodegenerative disease and age is the strongest risk factor, however sex ratio trends of the disease has opened up avenues of discussion on other causes as well (227-229). One of the most interesting example of exploring hypothesis through sex ratio trends was carried out by Morozova *et al* in their meta-analysis of age specific incidence of Parkinson's disease in different countries (226). The hypothesis that if smoking effectively reduces risk, Parkinson's disease should be relatively less common in populations that smoked more. According to authors, this hypothesis, however, was not directly testable because geographical and temporal variations in incidence of Parkinson's disease are often spurious, reflecting differences in diagnostic criteria, access to health care, and methods of case ascertainment rather than genuine variations in risk. In addition, they stated that differences among populations also reflected possible variation in underlying genetic susceptibility to Parkinson's disease. To circumvent this limitation, the investigators used the sex-ratio rather than absolute estimates of incidence of Parkinson's disease, because the male to female ratio in Parkinson's disease frequency should be relatively robust to variations in case ascertainment rates and to geographical differences in genetic susceptibility. For each country, and birth cohort, they estimated the sex ratio in incidence of Parkinson's disease, and correlated these ratios with corresponding sex ratios in smoking behavior. The analysis was similar to smoking and multiple sclerosis

sex ratio study by Palacios *et al* (32). The study authors concluded that results were consistent with the hypothesis that smoking reduces the risk of Parkinson's disease.

5.1.3 Sex ratio: incidence of rheumatic diseases

Many autoimmune diseases in the literature show a striking imbalance between males and females, with female predominance in most of these diseases (230). The male and female distribution of rheumatoid diseases vary considerably, in particular, rheumatoid arthritis which is a relatively common autoimmune inflammatory disease that affects the synovial joints, and with time due to the persistent inflammation of the joints causes significant functional loses. A high sex ratio (female-to-male) in rheumatoid arthritis (i.e., SR > 1) are usually attributed to the influence of estrogenic hormones, however Lockshin has criticized this view because the evidence that he reviewed demonstrates that the attributed sex ratios are imprecise because the definitions and classifications of autoimmune diseases differ substantially (231). Furthermore, these sex ratios were derived mostly from weak sources such as individual clinics, physician practices and voluntary agencies, and as such explains most of the variability in the sex ratio of incidence rheumatoid diseases (231-234). A study by Kvien *et al*., that used rheumatoid arthritis registry in Oslo, found that the sex ratio decreases after the age of 60 which presumably reflects gender-bias in referral practices by general practitioners (235). In fact they quote a Norwegian study that shows that women with rheumatoid arthritis are referred late compared to men (236). The sex ratio variability in rheumatoid diseases have however postulated on the environmental factors e.g., more men than women take drugs that induce lupus (low sex ratio), more men are exposed to silica inducers of scleroderma-like disease (low sex ratio), and more women have been exposed to the contaminated cooking oil that caused a scleroderma like illness in Spain (high sex ratio).

The epidemic of eosinophilia-myalgia syndrome was female predominant since more women than men took contaminated L-tryptophan, a putatively natural antidepressant; and the result was a high sex ratio (234). A review of animal model studies showed that the sex ratio is variable for rheumatoid diseases, and the causes for some of these diseases have remained mostly inconclusive (231).

5.1.4 Sex ratio: incidence of infectious diseases

It has been observed that, in general, males are more susceptible to infectious diseases than females. Multiple studies have documented this difference in susceptibility to certain infections both in humans and in a variety of other species (237-240). A recent meta-analysis on epidemiological trends of infectious disease by Guerra-Silveira and Abad-Franch attempted to explore two major, but not mutually exclusive hypotheses, to explain the different incidences in males and females (241). Using the disease incidence data from Brazilian Ministry of Health and Brazilian Institute of Geography and Statistics, the investigators explored the extent of physiological-sex-hypothesis versus behavioral-gender-hypothesis effects through the analysis of male relative to female incidence rates in several infectious diseases. The physiological hypothesis that they explored posited that the interactions between sex hormones and the immune system render one sex more susceptible to infection, with genetic differences likely playing some role, whereas the behavioral hypothesis posited that sex-biased infection rates emerge from sex-specific exposure to contagion e.g., differences in gender related behaviors. This study revealed a clear post pubertal high sex ratio (male-to-female) in the incidence of cutaneous leishmaniasis, tuberculosis, lepromatous leprosy, and leptospirosis. Infections, such as typhoid fever had no particular pattern whereas a low sex ratio pattern was found in dengue fever. Interestingly, a

high sex ratio in infectious disease susceptibility was observed in infancy, when behavior is supposed to be neutral, but where sex steroid levels transiently rise. Overall their findings contradict the behavioral hypothesis while matched with those of physiological hypothesis. The authors did point out to the limitations in the ascertainment of cases because of the uncertain quality of notification records. They believed, however, that the quality issues may not affect males and females differentially since they assessed the infections using relative measure (i.e., sex ratio) and that the results will hold if the ascertainment is equally poor in males and females. The finding from this study, however was critiqued by Markle and Fish who found the results provocative and that their results lacked a mechanistic approach at cellular level (237) and hence suggested more insight from rodent studies. Markle and Fish also pointed out that, rodent studies, can allow for dissection of sex-specific effects in immunity while controlling for exposures of pathogens.

Previously, sex ratios in reported incidences in infectious diseases of viral and bacterial origin among children (and adults) were examined in Israel over a period of twenty years (242). The observation that an increased susceptibility in male children to severity of some infectious diseases was the rationale for their study (although the overall rate of infection does not vary by sex in literature). Sex ratio analysis of incidence rates (males relative to females) of hepatitis A, meningitis, shigellosis, and salmonellosis was carried out to understand why there seems to be a male predominance in certain infections while not in others and why there is lack of uniformity in the literature. In the study, it was postulated that if the male predominance in infectious disease was due to a higher prevalence of those with increased susceptibility, such as relative immune deficiency, then failure to detect sex differences in the incidence of some diseases may be due to a statistical artifact resulting

from certain characteristics of the disease. Infectious diseases manifest with widely varying ratios of symptomatic to asymptomatic cases. Since only symptomatic cases are reported, it was postulated that this may be one of the possible reasons for the inconsistencies in the literature. A mathematical model was developed in an attempt to determine whether the variability in the symptomatic to asymptomatic case ratio may make the presence of sex differences in incidence more difficult to detect for certain diseases. One possible bias that was examined in the study was the completeness of ascertainment from ministry of health that reported infectious diseases based on a passive surveillance system, and therefore is far from complete. However, according to authors, there appears to be no reason to suspect selective differences in reporting for males and females since the analysis was based on relative measure.

5.1.5 Sex ratio: mortality of cardiovascular diseases

There are studies that have used sex ratio as a research tool for understanding causes of the differential mortality, in males and females, in several types of cardiovascular diseases such as ischemic heart disease, coronary heart disease, myocardial infarction and stroke (34, 35, 243-247). The investigators highlighted the temporal changes in the ratio for mortality rather than for incidence in assessing causes of changes (245). Zhang *et al*., undertook comparisons of SR for mortality between 27 populations from 1955 to 1990 to determine which causative factors (e.g., alcohol, animal fat, smoking, dietary sodium and potassium, fish consumption, and urinary cation excretion) influenced the SR (i.e., to ascertain whether any difference between countries could be attributed to a difference between men and women in the level of exposure to risk factors that affected mortality) (243, 245). Data on global stroke mortality was obtained from WHO, whereas risk factor data were obtained from

several sources such as Food and Agricultural Organization (FAO) and country wide epidemiological surveys. These ecological studies were an impetus for large scale studies, and substantiated several hypotheses later on, such as inverse association of fish consumption with lung cancer mortality in countries with high levels of smoking and fat consumption (248).

The earliest epidemiological study on sex ratio (male:female) of myocardial infarction was carried out by Lee and Thomas in 1955 (246) when they observed the incidence to be higher in women than men while also noticing the difference in two-time periods (i.e., 1910-1939 when sex ratio was one and in 1940-1954 when sex ratio was less than one). They used the data from autopsy and clinical record for the period 1910-1954. They hypothesized and explored three factors for this remarkable shift of sex ratio in two time periods: changes in body weight; weight of kidney as an index to the degree of systemic hypertension; and diabetes mellitus. They concluded that none of these factors were able to explain disproportionate rise in incidence among women in the latter time period.

Lawlor *et al*., examined secular trends and geographic variations in the mortality of coronary heart disease across fifty countries, and also investigated how the sex ratios relate to the distribution of risk factors (35). The sex ratio for mortality from coronary heart disease ranged from1.4 to 2.9 (males relative to females). The highest ratios were seen in Poland, France, and Norway and the lowest in rural China, Cuba, and Armenia. The sex ratio for mortality from coronary heart disease was found to be associated with mean per capita fat consumption. They also concluded that sex ratio in mortality varies over time and across countries in a way that could not be explained by *endogenous* estrogens and that these are likely to be driven by environmental factors. Artifacts were discussed such as coding

practices, diagnostic techniques, and increased life expectancy as possible explanations of sex ratio variations. The investigators, however believed changes in coding practices and increased survival should affect both men and women equally, and hence could not explain changes in sex ratios completely.

5.2 Sex ratio: a guide to interpret cancer incidence trends and hypotheses

Sex ratio in studies on cancer has previously been used to describe differences between males and females in the etiology and progression of tumors (38). Nicholson and Davis investigated changes in this ratio using mortality data from the US, Czechoslovakia, West Germany, Japan, and Italy (249). Cook *et al*., used a large dataset from the Surveillance, Epidemiology and End Results Program (SEER) in the US and computed age-adjusted male-to-female incidence rate ratios for specific cancers for the period 1975 to 2004 (37). Most cancers included in this review had higher incidence rates in males, which suggests the possibility of either universal mechanisms that increase male susceptibility to cancer or the uniqueness of men's occupational exposures that results in higher SR. Incidence rate ratios by sex were also explored by Edgren *et al*., to identify cancers in which there was a consistent gap between male and female incidence (39). Male-female incidence rate ratios were calculated for 35 cancers in 60 countries. The results showed that cancer incidence was statistically significantly higher in males than females for 32 cancer types with a male predominance of more than two-fold in 15 types and more than a four-fold increase in 5 types. The authors also carried out qualitative evaluation of SR by dividing the cancer types in three groups: (a) cancers where they deemed the sex differences enigmatic because different exposure to established risk factors was an inconceivable explanation; (b) cancers where they deemed that established risk factors at least partly explain the sex differences; and

(c) malignancies where they deemed that established risk factors provide a plausible explanation for the sex disparity. Through this qualitative assessment, the authors also observed that 13 cancers were entirely unexplained by smoking or alcohol.

5.2.1 Exogenous and endogenous causes of cancer

Recently, Radkiewicz *et al*., have discussed the intrinsic biological and environmental mechanisms through sex ratio of age adjusted incidence rates retrieved from Swedish cancer registry. These mechanisms have been divided into exogenous and endogenous factors associated with sex differences in cancer risk (250). Several authors in the past and present have provided an overview of *exogenous* and *endogenous* causes of cancers, thereby highlighting several interacting factors affecting risk of cancer (250-253). Greenwald has devoted a whole chapter for defining and understanding exogenous and endogenous causes of cancers alluding to the heterogenous nature of cancer (252). According to him, effective cancer prevention requires recognition of exogenous causes of the disease (i.e., factors that originate outside the body). Some experts, in fact, prefer to characterize these factors simply as *environmental*. The exogenous causes of cancer include agents and stimuli recognizable above the level of the cellular changes that initiate malignant disease. Phenomena of this kind are presumably more easily recognizable than those lurking in the deep recesses of cells. Once recognized, it is thought, they can be controlled. Physicians and scientists have recognized exogenous causes of cancer for hundreds of years. Discovery of such factors often began as an observation that some people, particularly those in certain trades or professions, have especially high risks of developing cancer. Exposure to hazards at the workplace, such as industrial chemicals, has presented the most obvious risk. Modern observers have included radiation, diet, and personal lifestyle practices in this area of

concern, expanding the notions of exogenous and environmental beyond their original meanings (252, 253). Cancers are also caused by *endogenous* factors (i.e., those that arise within the body, often at the cellular or molecular level). These factors are typically less open to direct observation and manipulation than exogenous factors e.g., age, endogenous hormones, genetics and heredity, and race (252). Lutz and Fekete have categorized these factors as avoidable (i.e., exogenous) and un-avoidable (i.e., endogenous) (253).

Hence in this chapter, several studies were presented that utilized sex ratios to generate or suggest some hypotheses regarding endogenous or exogenous causes of cancer by observing variation of sex ratios. Therefore, as suggested in several past and relatively recent studies (37-39, 249, 254), investigating changes in SR provides a useful guide for interpreting trends for the following reasons:

1) Comparison of SR can avoid problems that exist in interpreting incidence rates from cancer registries over time for males and females separately (e.g., changes in diagnostic techniques, tumor definitions, coding practices, errors in population bases in catchment areas of registries, and prevention strategies), since it is likely that these issues affect males and females to the same extent and are not likely to affect both sexes disproportionately (249). However, it should be kept in mind that SR can still be subject to the effects of gender differences in reporting behavior, health care access and utilization, and physician behavior in some cancer registries (37, 255, 256).

2) Analysis of time trends in SR can make the rate difference between males and females more evident, and can be more indicative of the presence of factors that may differentially affect either males or females. For example, hepatitis infections (type B and C) are more common in males than females and is likely to account for much of the rate

difference in liver cancer over time, but the anti-carcinogenic effect of estrogen can also be hypothesized for consistent lower rates of liver cancers in females (257). Another analogous example can be of smoking related cancers for which there is an indication that the prevalence of smoking has increased in females in some countries (258-261).

3) The extent to which environmental or health behaviors influence regional cancer rates may be similar in males and females and, if so, the effect of such changes is lessened when observing SR (in contrast to specific rates for each sex separately) (249). In some cancers, the changes over time can be substantial and exhibits differences in males and females that are sufficiently large to warrant further investigation e.g., occupational exposures that are shown to be strongly determined by sex (262).

4) Spatial correlations of SR (correlations between adjacent geographic regions) for cancer types can provide some indication whether the causes of those cancers have some pattern e.g., clustered, random, or dispersed patterns and thus provide some evidence to inform hypotheses of endogenous versus exogenous causes.

An important point to be noted here is, while we are assuming similar ascertainment in males and females, there is a possibility that the ascertainment could also be gender-driven specifically in low-resource registries. For this reason, it is also important to examine the contextual problem of potential gender-bias in the cancer registries.

Chapter 6 Rationale and Objective

6.1 Rationale

In the preceding chapters, an account is narrated that comparisons of incidence and mortality rates of cancers are carried out across time and geography, and that these comparisons are affected by several artifacts. These rates are retrieved through several sources, either cancer registry for incidence, or vital registration system such as death certificates for mortality. There are roughly 600 cancer registries around the world that are divided into national and regional cancer registries from low-, middle-, and high-income countries. These registries are resource dependent and they use different methods of cancer case ascertainment. CI5 provides data on cancer incidence from wide geographic regions, while adhering to the various quality criteria and uses standardized methods of completeness for ascertainment. In spite of stringent quality criteria maintained by CI5 for any cancer registry to be included in its database, the data might not be always comparable across registries in low and high resource registries because of several artifacts in play.

International comparisons in cancer incidence are potentially important sources of evidence for generating hypotheses about cancer etiology. However, the estimates of inter-regional variation in cancer incidence is compromised by imperfect accuracy and completeness of available cancer incidence data. Such imperfection can result from artifacts such as problems of access to quality medical diagnostic services or from inadequacies in the cancer registration process. Since these artifacts of cancer diagnosis and registration can operate differentially in different regions and over different time periods, comparisons of cancer incidence across regions or between time periods can be biased and misleading. As an alternative to comparing cancer incidence between regions or time periods, we propose that

the SR of type-specific incidence (male incidence/female incidence) is a more stable parameter, less susceptible to distortion than the male and female rates on which it is based. This holds to the extent that, within any region or time period, errors and incompleteness of cancer diagnosis and registration apply equally to both males and females. Assuming approximate validity of the estimated SR in cancer incidence, we can describe and assess the geographic and temporal comparisons in estimated SRs and assume that these accurately reflect the geographic and temporal comparisons in true SR. Just as ecologic comparisons (i.e. geographic or temporal) in disease incidence can be a fruitful basis for developing or confirming hypotheses on the etiology of disease, so can ecologic comparisons (i.e. geographic or temporal) in SR. This thesis is designed to derive geographic and temporal comparisons in the SR for each type of cancer, by computing the variability of SR and then estimating the geographic and temporal trends in SR variability for each cancer. Such an information base will be a useful ground for etiologic speculation.

6.2 Objectives

This thesis examine variability in the ratio of male to female incidence rates (i.e. the SR) for different cancer types across large number of world-wide cancer registries. This exercise is carried out to enable cancer epidemiologists and future investigators to consider sex ratios when comparing incidence rates across time and geographic boundaries. The reason is to facilitate them in generating hypotheses and to reflect on to the inferences that can thereby be drawn concerning possible etiologic influences. Since a contextual problem of gender-bias can mask some of the inferences, hence it is important to also explore gender-bias potential or gender inequalities that can exists in access to health care and cancer diagnosis and treatment on selected cancer types.

Therefore, the main aim of this study is:

- To infer as to potential causes that drive sex ratio variability (i.e., the ratio of male to female incidence rates), of type specific cancers across time and geography, generating hypotheses.

The secondary aim is:

- To explore the extent to which country-level gender inequalities can provide clues on quality of cancer registries for selected cancer types through sex ratios.

Chapter 7 Methodology

7.1 Overview of design

To describe variability in SR of cancer incidence across time and region for each type of cancer, we first assembled a comprehensive set of published incidence rates for males and females, from large number of cancer registries and across many years of observation (A total of 77 in 1973-77; 142 in 1988-92; 281 in 2003-07 and 113 longest surviving registries from 1953 to 2007). For each registry and each time period and each type of cancer (30 types of cancers in each of three time periods of 1973-77; 1988-92; 2003-07 and 28 cancer types in 1953-2007), we then computed the ratio of incidence among males to incidence among females, termed the *SR*. For each type of cancer, we then conducted analyses of SR to estimate the following parameters:

1. For each of selected time period, the *magnitude* of SR (SRm) and *variability* in SR (SRv) across registries worldwide.

2. *Temporal trends in* SR in the longest surviving registries, as well as estimates of *geographic variability* (*inter- and intra-registry regional variability,* using registry as a synonym for a specific region in which the registry itself is situated) in SR.

3. For the most recent time period (2003-07), the *spatial autocorrelation* in SRs across registries in Europe and Asia.

Descriptive methods are used primarily in the thesis, with recourse to mixed-effect regression methods particularly for studying temporal trends, and geographic variation of SR. Mixed-effects regression was also used to examine gender inequality across some the cancer

registries for selected types of cancers. A method known as Moran's Index is used for estimating spatial autocorrelations of SR.

The results generated are used to describe and characterize the various cancer types as embodying low or high degrees of variability in SRs. In general, while the magnitude of sex ratio is defined in terms of low (SR < 1.0, indicating higher risk in females) or high (SR > 1.0, indicating higher risk in males), both magnitude and its variability are also defined according to the tertiles of values obtained for total number of cancers (30 cancer types in this study). The inferences are made later in Chapter 8 on results from these categorization, and hence the term "low" or "high" sex ratios (SRm or SRv) in the thesis is used in multiple places. The terms "low" and "high" SR are also used for a specific type of cancer when comparison is made across different cancer registries e.g., lung cancer SR might be low in Sweden compared to Egypt where it might be high, when in both registries SR is greater than value of 1 (i.e., higher risk in males in general).

In the study, registries from CI5Plus data were considered as "regions". These registries are specific to counties, provinces, prefectures, states, capital cities, and national registries from an individual country. The inter- and intra- regional variabilities of SR were computed using the CI5Plus data on incidence rates from these 113 cancer registries. Throughout the thesis, these registries are referred to as "regions". Figure 7-1 also shows the timeline of these registries. The inter- and intra- regional variabilities of SR for each cancer is computed through mixed-effect regression modelling, to show which type of variability can be potentially important for a cancer type, and to posit potential questions and answers when variations in intra- compared to inter- is larger, smaller, or similar. Inter-regional variability is defined based on the trajectories (slopes in regression model) specific to different registry-

region, and specific to a cancer type, as "*variation in sex ratios between different registry-regions, at the same point in time or calendar year*". The model also adjusts intra-regional variability of SR, which is defined as the "*variability of SR within registry-regions over a period of time"*.

The spatial patterns of SR in two continents (Asia and Europe) are used to rank each cancer type according to degree of clustering from extremely high to low. All this information is used to gain insight into possible avenues for exploration of the etiologies of the different cancer types.

Cancer types for which SR variation was found to be high and where tobacco-smoking could provide a plausible explanation for sex-disparity, we compared selected cancer registries from countries that were presumed to be high and low in terms of gender inequalities with country-level prevalence of smoking among males and females.

To infer from results of analyses, on causes of cancers, the terms, "endogenous" and "exogenous" were used. The definitions of these terminologies are based on Lutz and Fekete (253) who categorized these causes as avoidable causes, i.e., the factors that originate outside the body and characterized as environmental in origin including diet, radiation, and personal lifestyle practices. Endogenous causes or un-avoidable causes are factors that cannot be manipulated like exogenous factors e.g., age, hormones, genetic and heredity, and race.

7.2 Data sources

Cancer Incidence in Five Continents (CI-5) now comprises ten volumes covering cancer incidence reports from many cancer registries over the past 60 years (145). The published volumes of CI-5 include tabulations of cancer incidence rates in two basic formats:

1. Registry-specific tables showing incidence rates according to sex, age group and type of cancer.

2. Tables of summary rates (crude incidence rates and age-standardized incidence rates for each cancer type).

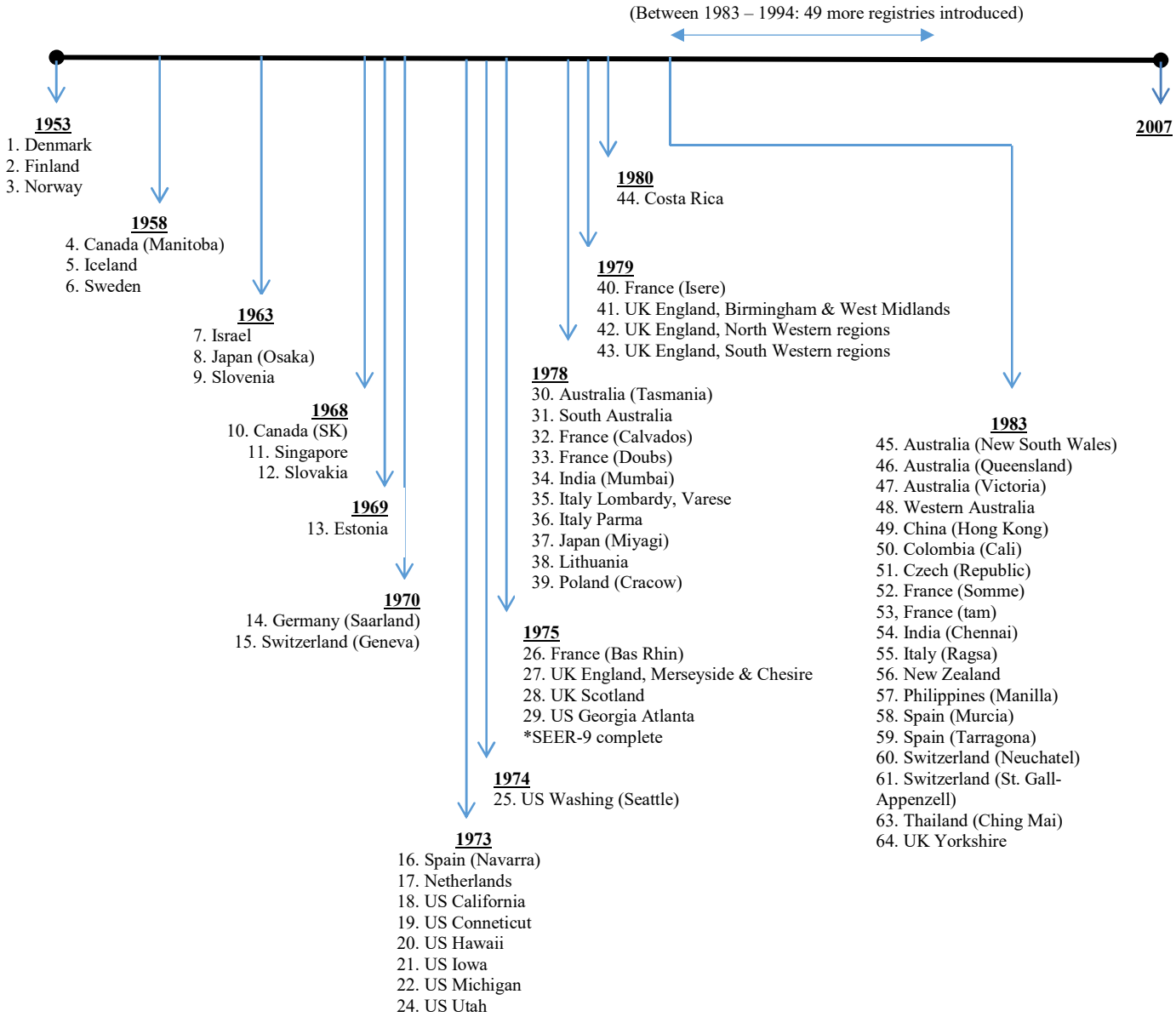
The CI5*plus* database contains updated *annual* incidence rates for 113 cancer registries published in CI5, for the longest period available from 1953 up to 2007 (see Figure 7-1), for 28 major types of cancers. Both CI5 volumes I-X and CI-5plus comprise a public domain website, accessible at www.ci5.iarc.fr. To carry out the analyses listed in Section 7.1, the following CI-5 databases were accessed from IARC website:

1. For the analysis of the *magnitude* of the SR across registries and *variability in* SR across registries*:* CI-5 Volume IV (1973-1977); CI-5 Volume VII (1988-1992); CI-5 Volume X (2003-2007).
2. For the analysis of time trends in SR and inter- and intra- regional variability in SR: CI-5plus (1953-2007).
3. For the spatial autocorrelation in SR across regions represented by registries in Europe and Asia: CI-5 Volume X (2003-2007).
4. The Gender Inequality Index (GII) Database of United Nations was also accessed for the year 2005 in order to select and compare registries of countries that have high and low GII (49). The selected registries belonged to recent volume of CI5 (2003-07). The age standardized prevalence estimates of current smoking were retrieved from WHO Report on global tobacco epidemic of current smokers published in 2005. Prevalence estimates were for adults aged 15 years and over, resulting from analysis of adult tobacco surveys completed by countries since 1990 and reported in 2003. Tobacco smoking includes

cigarettes, cigars, pipes or any other smoked tobacco products. Current smoking includes both daily and non-daily smoking.

Figure 7-1: Timeline of 113 long standing cancer registries and year they started operating in CI5Plus database.

First three registries started in 1953 and reported incidence rates every year upto 2007. Subsequent registries worldwide followed in different years. No new registries were introduced after 1994.



7.3 Computation of incidence rates

Because of differing age distributions in different populations covered by CI5, it is important to ensure that all comparisons of incidence rates or parameters derived from incidence rates, are based on age-standardized rates. Each CI5 volume provides age-standardized incidence rates (ASIR) for each cancer type in males and in females in each registry covered by the volumes.

The crude incidence rate (CIR) is the rate at which new cases occur in a population during a specific period. This rate is classically expressed as the average number of cases occurring per 100 000 persons each year or 100 000 person-years. It is computed with the following formula:

$$CIR = \frac{Number\ of\ new\ cancer\ cases\ observed\ in\ the\ period}{Total\ population\ in\ the\ period}$$

The ASIR is expressed, as is the CIR, as the number of new cases per 100 000 person-years. The calculation is a weighted average of age-specific rates:

$$ASIR = \sum_i \frac{di\ wi}{yi} \quad \text{..... } \textit{(Equation 1)}$$

Such that $\boldsymbol{i}$ represents each age group, $\boldsymbol{d_i}$ the number of cases in the *ith group*, $\boldsymbol{y_i}$ the population size in the *ith group*, with $\boldsymbol{d_i/y_i}$ being the age specific rates for each *ith category* and the sum of w_i (weight from standard population) being equal to 100 000 to express the age-standardized rate per 100 000 person-years (see also Appendix 1).

The standard population used by CI5 is one that was initially proposed by Segi in 1960, and subsequently modified by Doll et al (110). It was based on the age structure of the

pooling of populations from 46 countries. An example of the impact of age standardization is the following: in Denmark in 1998-2002 the crude incidence rate of melanoma among males was 17.46 cases per 100 000 person-years. Using the standard population that is used by CI5, the age-standardized rate was 11.94 per 100 000 person-years. The age-standardized rate is lower than the crude rate because the standard population was on average younger than the Danish population. However, the age-standardized rate of 11.94 can be compared with other rates standardized on the world population (see also Appendix 1).

7.3.1 Population at risk

Population at risk was used as the denominator in the formulas for calculating crude or adjusted incidence rates. These denominators are often disregarded, and users tend to focus on incidence rates and discuss variability in incidence data gathering or ascertainment (e.g., completeness, classification). This notion is driven by the age standardization process, which aims to eliminate the effect of population age structure as a confounding factor in comparing rates across populations. Numbers of persons at risk are routinely collected by registries from official statistics offices for example, and registries provide these numbers to the Cancer Incidence in Five Continents, along with incidence rates.

7.4 Sex Ratio definition and computation

Based on computations of ASIR, we defined SR as ASIR in males relative to ASIR in females for each cancer type specific to individual registries as well as time-period (s):

$$SR_{ij}^{k} = \frac{ASIR_{ij}^{k}(male)}{ASIR_{ij}^{k}(female)}, \ldots\ldots\ldots\ldots\ldots\ldots\ldots \textit{(Equation2)}$$

where i : time period and $i = 1,\ldots,N$

j : registry and $j = 1,\ldots,n_i$

k: cancer type and $k = 1,\ldots,K$

7.5 Estimating Sex Ratio magnitude (SRm) and variance (SRv) across registries

7.5.1 Data extraction

The data files from ten volumes of CI-5 are available from IARC websites http://ci5.iarc.fr/CI5i-ix/ci5i-ix.htm in exportable format (.txt format). The number of cancer cases of different types in all age groups of 5 year intervals (0-4, 5-9, 10-14… 80-84, 85+) were obtained for males and females from cancer registries across the world.

Crude incidence rates (CIR) and age-standardized incidence rates (ASIR) per 100,000 are available in three formats from CI5: *by populations (registries)*, *by cancer type*, and *by volumes* signifying time periods. The option of *'by population'* presents CIR and ASIR by registry for specific cancers and sex. Therefore, we were able to extract data files in multiple small steps for all populations in all age groups for each cancer, separately for males and females. The second format of data retrieval was *by cancer type.* For analyzing time-periods, information was collected using the *volume* format.

All information retrieved from IARC-CI5 data files was stored using a statistical software package (SPSS version 23.0). While data were retrieved from all ten volumes of CI-5, only three volumes were selected to be cleaned, coded, and analyzed: CI-5 volumes IV, VII, and X.

7.5.2 Estimating in three time periods (1973-77; 1988-92; 2003-07)

Although it was possible to assess variability in SRs across registries in any of the ten periods corresponding to the ten volumes of CI5, we chose to assess it in each of three time periods, corresponding to volumes IV (1973-1977), VII (1988-1992), and X (2003-2007) of CI5. These were chosen to permit the maximum number of registries spanning a large number of geographic regions, with a gap of ten years between each time period to allow for observation of time-trends in SR. Moreover cancer coding techniques based on ICD-10-O were compatible across the three time periods.

For each cancer, in each of the three time periods, we computed two parameters related to the SR, namely the *magnitude of the SR* and the *variation in the SR* across registries. To facilitate presentation, we categorized each parameter into three levels (i.e., low, medium, high). The purpose of creating these two dimensions and three levels was to permit speculation on the patterns and causes of cancers. Also, it allowed us to ascertain whether there are uniform or unstable patterns of cancer types in the three time periods, and whether the patterns hold if SRm and SRv are stratified according to low, medium and high incidence rates of cancer.

7.5.3 SRm: definition and computation

The ASIRs in males relative to females for each cancer type from each registry-population were averaged across all available registries. The *average value* of the SR for a

particular cancer in a given year is termed "the magnitude of the SR (*SRm*)". Based on equation 2, Table 7-1 shows calculation of *SRm* denoted as SRm_i^k in 281 cancer registries in the time period 2003-07 using the following formula: $\mu = (\sum X_j)/ 281$, where $X_j := SR_{ij}^1$ is the SR for a registry population '*j*' in selected time period and cancer type (column a). The mean SRm was then calculated using the formula:

$$SRm_i^k = \mu = \sum_{j=1}^{n_i} SR_{ij}^k \ / \ n_i \ \ldots\ldots\ldots\ldots\ldots\ldots \textit{(Equation 3}$$

7.5.4 SRv: definition and computation

For a specific cancer and time period, *SRv* was calculated by taking the average of the squared differences of the SRm across a total of 281 registries. Table 7-1 describes the calculation of *SRv* denoted by SRv_i^k using the following formula:

$$SRv_i^k = \delta^2 \ = \sum_{j=1}^{n_i} (SR_{ij}^k - \mu)^2 / \ n_i \ \ldots\ldots\ldots\ldots\ldots\ldots \textit{(Equation 4)}$$

SRm and SRv were then stratified according to low, medium, and high incidence. ASIRs for each cancer from each registry-year were summed across the total number of registries and averaged. The average value of ASIR for a specific cancer is termed the magnitude of incidence of that cancer (*Im*).

All computations in Sections 7.5.3 and 7.5.4 were carried out for all three volumes of CI-5 for periods: 2003-07, 1988-1992, and 1973-77, as well as for log transformation in section 7.5.5.

Table 7-1: Magnitude of the mean incident SR (SRm_i^k) and the variance of the incident SR (SRv_i^k) for 30 cancer types across cancer registries in a selected CI5 volume (CI5-X: 2003-07*).

Sex Ratios of cancer incidence worldwide in a given time period (denoted i)

Cancer registries ($n_i = 281$)	***Cancer types (K = 30)***			
	SR_{ij}^1 : ***Bladder***		SR_{ij}^{30} : ***Tongue***	
	(a) $X_j := SR_{ij}^1$	(b) $(X_j - \mu)^2$	(a) $X_j := SR_{ij}^{30}$	(b) $(X_j - \mu)^2$
SR_{i1}^k : ***Algeria***	X_1	$(X_1 - \mu)^2$	X_1	$(X_1 - \mu)^2$
SR_{i2}^k : ***Argentina***	X_2	$(X_2 - \mu)^2$	X_2	$(X_2 - \mu)^2$
SR_{i3}^k : ***Australia***	X_3	$(X_3 - \mu)^2$	X_3	$(X_3 - \mu)^2$
⋮	⋮	⋮	⋮	⋮
SR_{i281}^k : ***Zimbabwe***	X_{281}	$(X_{281} - \mu)^2$	X_{281}	$(X_{281} - \mu)^2$
	$\mu = (\sum X_j)/281$	$\delta^2 = \sum(X_j - \mu)^2/281$	$\mu = (\sum X_j)/281$	$\delta^2 = \sum(X_j - \mu)^2/281$

$$\boldsymbol{SRm_i^k} = \mu = \sum_{j=1}^{n_i} SR_{ij}^k \;/\; n_i \qquad \boldsymbol{SRv_i^k} = \sum_{j=1}^{n_i} (SR_{ij}^k - \mu)^2 / n_i$$

**Computations were repeated on 77 and 142 cancer registries in CI5 volume IV (1973-77) and volume VII (1988-92).*

7.5.5 Log transformation of SRm and SRv

Log transformation was carried out in a similar way as shown in Table 7-1. The mean and variance of log values of SRs were computed (for logs of each $X_1, X_2 \ldots X_n$ and then mean and variance was computed). Hence the log transformed versions of equation 3 and 4 are presented in equations 5 and 6 as follows:

$$lSRm_i^k = \mu_l = \frac{\sum_{j=1}^{n_i} \log(SR_{ij}^k)}{n_i}$$

…….…… *(Equation 5)*

$$lSRv_i^k = \sigma_l^2 = \frac{\sum_{j=1}^{n_i} (\log(SR_{ij}^k) - \mu_l)^2}{n_i}$$

…… *(Equation 6)*

7.5.6 Generating SRm × SRv Tables

All registries in the CI-5 volumes were coded and statistics were generated including the number of cases for each cancer type, the mean, variance, and standard deviations of the SR. Tertiles were used to categorize the SR mean and variance (see Appendix 8 for details), thereby enabling creation of matrices of SRm and SRv (Figure 7-2) as well as stratification of these matrices according to low, medium and high levels of incidence rates (Figures 7-3 a, b and c). Besides computing each SRm and SRv for each cancer, the top three registries with the highest and lowest SRm were also tabulated. Descriptive statistics were computed by taking the mean and variance of the log values to generate *log-transformed* versions of SRm and SRv (See also Appendix 9, pages 280-81)

Figure 7-2: Levels of magnitude and variances of SR (*SRm*×*SRv*).

Different types of cancers are distributed in each of nine cells according to low (L), medium (M), and high (H) levels of SRm and SRv generated through tertiles

SRm \ SRv	L	M	H
L	1	2	3
M	4	5	6
H	7	8	9

Figure 7-3: Levels of sex ratio magnitude (SRm) and variance (SRv) stratified according to levels of magnitude of cancer incidence (a) low: Im_1 (b) medium: Im_2 (c) high: Im_3

Im_1

SRm \ SRv	L	M	H
L	1	2	3
M	4	5	6
H	7	8	9

(a)

Im_2

SRm \ SRv	L	M	H
L	1	2	3
M	4	5	6
H	7	8	9

(b)

Im_3

SRm \ SRv	L	M	H
L	1	2	3
M	4	5	6
H	7	8	9

(c)

7.7 Estimating time trends and geographic variability in Sex Ratios

7.7.1 Data extraction

The analysis of time trends in SR, as well as between-registry and within-registry (inter- and intra-regional) variability in SR was carried out for long-standing cancer registries from 1953-2007 for 21 different types of cancers. The CI-5plus database was accessed to carry out the analysis. This database was also used to test whether gender inequality can predict changes in sex ratios across time for selected cancer types and registries. The earliest cancer registries in this database started functioning in 1953 in Denmark, Finland, and Norway

(Figure 7-1). Afterwards, other cancer registries were initiated in different years. Once initiated, these registries reported cancer incidence every year until 2007 (the last reported year in most recent volume of CI5). No new registry was introduced after 1994 in this database.

7.7.2 Data setup

The introduction of cancer registries with different starting points from 1953 to 1994 created an unbalanced dataset where different registries had different total numbers of observations on SR in each given calendar year. Therefore, the data from CI-5plus were extracted and organized so as to permit analysis of temporal trends in SR and inter- and intra-regional variability in SR in these different situations: *Scenario A:* analysis of all cancer registries that reported incidence for any year from 1953 to 2007 (i.e., all available observations from this period were analyzed); *Scenario B:* analysis of all cancer registries that reported incidence in any year from 1983 to 2007 including some registries that reported for part of that period (i.e., all available observations from this period were analyzed); and *Scenario C*: analysis of cancer registries that reported incidence every year from 1983 to 2007 (i.e., all observations in 1983 were exclusively analyzed until 2007). The details of these data scenarios are as follows:

(1) Scenario A: unbalanced registry data from 1953 to 2007:

Scenario A consists of data on observations of SR in different registries from 1953 to 2007. Different years can have different number of registries 'N'. These 'Ns' increase monotonically over time as new registries were introduced e.g., 1953 (N=3); 1958 (N=6); 1963 (N=9); 1968 (N=14); 1970 (N=16); 1973 (N=31); 1974 (N=32); 1975 (N=37); 1978 (N=48); 1979 (N=52); 1980 (N=53); 1983 (N=76); 1985 (N=81); 1989 (N=95); 1993

(n=112); and 1994 (N=113). In short, Scenario A begins with 3 registries in 1953 and ends with 113 registries reporting incidence rates upto 2007. These are called *unbalanced data* because different years can have different 'Ns'.

(2) Scenario B: unbalanced registry data from 1983 to 2007.

This extraction consists of data on SR from 1983 to 2007 (i.e., all the data before 1983 is removed in Scenario B). The minimum number of registries 'N' in this scenario is 76 and maximum is 113 (hence it is still called *unbalanced data*). The year 1983 was chosen as a cut point for this analysis (and the next scenario C) because N in this year (i.e., 76) is substantial enough sample size to facilitate in regression model building for 25 years' time period. A smaller 'N' or few years of repeated observations could become problematic in the analysis of inter- and intra- regional variability of SR (263).

(3) Scenario C: balanced registry data from 1983 to 2007.

This scenario consists of data from 1983 to 2007 and restricted to those registries (N=76) that were functioning and reporting throughout the period 1983 to 2007 (thereby resulting in balanced data). Any registry that joined after 1983 was removed from the analysis in this scenario.

7.7.3 Description of regression models in different data scenarios

Different analytic strategies were used depending on the type of data scenario from the CI5Plus database. After specifying three scenarios in Section 7.6.2, we next introduce four regression models for each scenario (and for each cancer type). The availability of incidence rates from the varying number of registries, with different ranges of time, allowed us to interpret results from different regression models, and choose one that performs relatively

better, with minimum error. Moreover, the SR is recorded as a repeated measure (i.e., SR is measured every year for a specific registry and as such the yearly measurements are not independent). With data such as these, an appropriate analytic strategy is the mixed-effects regression technique. However, this will require detailed step-by-step model-building as described below.

7.7.3.1 Setting up mixed-effects models

The strategy of building mixed-effects regression models in the current study (where the outcome variable is SR) is based on recommendations by Field (263) and Twisk (264): to start with a 'basic linear model' in which parameters are fixed, and then add random coefficients as appropriate. One advantage of doing this is that one can simultaneously compare the fit of the models as one adds in random effects parameters (263).

We first introduce two basic regression models (*a* and *b*), by adding independent variables (Appendix 2) specifically *'calendar year'* (model *a*) and then adding *'registry'* to model *a* (model *b*). Both of these basic regression models are *Fixed-Effect models* because they have fixed-coefficients (i.e., constant intercepts (β_0: average value of SR at baseline year) with constant regression slopes (β_1)). Therefore, model *a* has a fixed effect of 'calendar year' which is assumed to be same for every registry, whereas model *b* has a fixed effect of 'registry'. These models assume that the year effect is the same across all registries and the registry effect is the same for all years. They also assume that the temporal trend is linear. While these assumptions are simplistic and likely incorrect, they provide the baseline against which to judge more sophisticated models.

Next, for model *c*, we allowed the intercept β_0 to assume different value across different registries (Appendix 2). Because the intercepts (i.e., SR in the baseline year) are considered

to be random draws from a normal distribution, this model is labelled a *'Random-intercepts model'* (RI model). This model allows estimation of intra- and inter-regional variation of SR [designated $(\delta^k)^2$ and $(\tau^k)^2$ respectively].

The fourth regression model *d* is extended from model c (Appendix 2). Here, we not only allow SR at baseline year to vary for each registry, but also allow the slope of each registry to vary. This is called a *'Random-intercepts & Random-slopes model'* (RI and RS model). Due to the greater flexibility of this model, it can provide a more realistic estimation of intra- and inter- regional variation of SR [designated $(\delta^k)^2$ and $(\tau_1^k)^2$ respectively].

Both models *c* and *d* are called mixed-effects models because they include both fixed and random parameters. SPSS version 23.0 is used to store as well as analyze the data through a series of regression models. The interpretations of basic linear models (*a* and *b*) and mixed-effects models (*c* and *d*) in each of three scenarios are described as Appendix 2.

Limitations of fixed-effect models:

Fixed-effects models make the assumption that errors, including errors within individual registries, are independent. In our data, we believe that the SR observations are clustered within years for each registry. Clearly, SR in one year for a specific registry is related to the next year and so on. In other words, each SR observation for each year is derived from the same registry (e.g., SR in the cancer registry for Sweden in 1954 is likely to be related to the SR in 1955, 1956, etc.). Therefore, these observations are not independent.

In model *b*, the addition of individual registries as fixed effects greatly increases the number of parameters (β_n^k for each registry in the model b: Appendix 2). There are therefore

many parallel regression lines, one for each registry. Due to the large number of parameters β_n^k, this model has the potential to overfit the data.

7.7.3.2 Choice of mixed-effects models' type

Mixed-effects regression analysis provides a more flexible solution to the problems expected in fixed-effects models. First, non-independence in the current data provides justification to carry out this type of modelling. Second, overfitting can be avoided by introducing random coefficients as opposed to fixed. And finally, this type of modelling allows us to investigate the inter- and intra-regional variability of SR (263).

We are interested in assessing the overall effect of time (years) on SR while taking inter-regional variation into account. We allow for this using random intercepts (model c) and random slopes (model d) specific to each regional registry.

Random-intercept model (RI model):

The random-intercept model *c* is doing the same thing as the fixed-effect model *b*, but instead of having a separate intercept parameter for each registry, we now assume that the different study-specific intercepts are randomly and normally distributed.

Random-intercept plus random-slopes model (RI & RS model):

In this model *d,* we introduce the randomness of slopes in addition to randomness of intercepts. This means that the rate of change of the SR is allowed to vary randomly between registries. Following is the model used (see also Appendix 3 for comparison of interpretations with other models *a, b* and *c*):

$$SR_{ij}^{k} = \beta_0^k + \beta_1^k year_i + u_{1j}^k + u_{2j}^k(year_i) + \varepsilon_{ij}^k, \quad where\ \varepsilon_{ij}^k \sim N\,(0, (\delta^k)^2)$$

…… *(Equation 5)*

$k =$ cancer type; $i =$ index of calender year; $j =$ registry

$\varepsilon_{ij}^k \sim N\,(0, (\delta^k)^2)$ *is* the assumption that the errors are normally and independantly distributed with mean 0 *a*nd variance $(\delta^k)^2$

and $u_{1j}^k \sim N\left(0, \tau_1^{k2}\right)$ for registry *j* ;

$u_{2j}^k \sim N\left(0, \tau_2^{k2}\right)$ for registry *j* ;

$Cov\left(u_{1j}^k, u_{2j}^k\right) = \tau_{12}^k$

u_{1j}^k is the random variations of intercepts of SR among registries and it has a variance τ_1^{k2}; and u_{2j}^k is the random variations of slopes and it has a variance τ_2^{k2}

The random intercepts and slopes are correlated $Cov\left(u_{1j}^k, u_{2j}^k\right) = \tau_{12}^k$

[The results from RI & RS Models in each of the three scenarios are summarized in the next chapter (Tables 8-12 to 8-14)].

7.7.3.3 Assessing the fit and comparisons of models

The details of analysis are in Appendix 6 (A 6-a to A 6-u). In all Tables, the values of AIC (Akaike Information Criteria) are given for each mixed-effect model *c* and *d* in the column labelled 'Errors and AIC'. AIC is basically a goodness-of-fit measure that is corrected for model complexity and lower values mean a better fitting model. AIC is also known as the adjusted version of log-likelihood that is also used to assess model fit (263). In all models of mixed-effect regression analysis, the choice of estimation method to obtain better estimates of all parameters was the maximum likelihood (ML). Had the models used more predictors (independent variables) with multiple levels (multi-level models), the choice could have been restricted maximum likelihood (REML). This estimation method, REML,

only measures fit of the random components in the model: the fit of random effects could become better with fixed effect of time and other potential predictors, because then, we could be better defining the errors/residuals in the model by having more accurate fixed effects. Hence, the preferred estimation method in the study was ML and not REML.

In summary, we obtained changes in SR over time along with inter- and intra- regional variation of SR, by first dealing with non-independence of SR observations among different registries by adding random-effect (s). But before specifying random effect models, a fixed effect model was constructed where there was an average effect of time (individual years of incidence reporting) for all registries [i.e., when there was no assumption of random effects, neither of registry u_{1j}^{k} nor of calendar year with registry $u_{2j}^{k}(year_i)$ such as in RI & RS models]. However, in the study data, the residuals of a particular registry (the distance to an overall fitted line) are all correlated, i.e., those 25 residuals representing SR for each year from 1983 to 2007 for a single registry are not independent. In other words, SR in the first year in a particular registry is related to the second year and then to the third and so on. This issue was taken care of while systematically transitioning from fixed-effect models to RI-model, and then finally to RI & RS model, by specifying covariance structure of each registries' residuals in the data. There are different types of covariance structures that one can use, and each type of structure simply specifies a form of the variance-covariance matrix (a matrix in which diagonal elements are variances and off-diagonal elements are covariances) (263). For RI-model (model *c*), the type that was used was *Variance Component* (VC) which is a very simple covariance structure and assumes that all random effects are independent. For our RI & RS model, the choice was *Unstructured-Covariance structure* (UN) where every covariance and variance was considered unique, and was uniquely estimated in the data

(for RI & RS model, we wanted unique estimates for variances of intercepts, slopes and covariance of intercepts and slopes). Except for a few cancers, the AIC of the mixed-effect models showed that overall the RI & RS model was a better fit model and improved with the choice of variance structure.

7.7.3.4 Incorporating gender inequality in the analyses

We combined the country-level data from the United Nation's gender inequality index (GII), the Cancer Incidence in Five Continents (CI5plus), and The World Health Organization's report on Global Tobacco Health Survey (265). The country-level smoking prevalence data was obtained from this report on global tobacco epidemic. Overall, the gender inequality index reflects how women are disadvantaged in three dimensions (reproductive health, empowerment and labor market) and this data was retrieved for year 2005. The values of this index range between 0 and 1, with higher index indicating higher inequalities and thus higher loss to human development. Since for certain cancer types, smoking tobacco is an established risk factor, we examined the age standardized prevalence estimates of smoking and SRs for smoking-associated cancers in registries representing some of those countries that rank low and high on gender inequality index. Smokers were defined as those who smoked everyday any tobacco product such as cigarettes, cigars, pipes, bidis, etc (at the time of surveys in 2003 and before). For cancers such as of lung, bladder, esophagus, and larynx, we used unexpected deviance (e.g., large gender inequality and differences in smoking prevalence among men and women, but low sex ratio to identify cancer registries with possible gender biases in data quality.

Equation 5 (page 100) was modified to incorporate gender inequality in mixed-effects model. In this mixed-effects model with random intercept plus slope model, we included

high and low categories of gender inequality (from 32 countries) to test whether gender inequality has an impact on sex ratio changes over calendar years (1983-2007). In addition to the linear trend of time (in years), quadratic trend was also included. Furthermore, gender inequality and its interaction with time (in years) was also included in the model.

7.8 Estimating spatial auto-correlations of SR

In geographic health research, spatial auto-correlations are often computed to investigate variables that are correlated between adjacent locations (254, 266). In common statistical approaches, it is often assumed that measured outcomes are independent of each other (such as in fixed-effects models and between registries in the mixed-effect models that we used). In spatial data, it is often the case that some or all outcome measures exhibit spatial auto-correlation. An example of spatial auto-correlation is when the outcomes of two geographic points are inversely related to their distance apart.

7.8.1 Analyzing spatial autocorrelation: Moran's Index.

The first law of geography states that *"everything is related to everything else, but near things are more related than distant things"* (82). In statistical terms, positive spatial autocorrelation indicates that neighboring values are similar, suggesting spatial dependency; negative spatial autocorrelation indicates that neighboring values are dissimilar, suggesting inverse spatial dependence. An autocorrelation value of 0 implies that there is no spatial pattern. While several methods of measuring spatial autocorrelation have been explored in geological, ecological, health, and environmental studies; one of the most commonly used methods remains the Moran's Index (I) statistic (267, 268). The computation of Moran's I is based on populations whereas other methods (e.g., Geary's coefficient) are based on samples (269). Since in this study, population-based cancer registries are used, the global Moran's I is

the most appropriate method to explore the spatial patterns of incidence rates and SR of incidence rates for different types of cancers. We restrict our analyses to regions of Europe and Asia which are connected by land mass.

Global Moran's I measures the spatial autocorrelation of *feature locations* (i.e., registries representing cities/regions in Europe and Asia in this study) and *feature attributes or values* (i.e., SR of cancer incidence rate) simultaneously. To explore the overall spatial patterns of the most common cancers in Europe and Asia, the global Moran's I statistic was used to represent the degree of clustering. It is calculated through a standard formula as follows:

$$I = \frac{n \sum i \ \sum j \ W_{ij}(X_i - \bar{X})(X_j - \bar{X})}{\sum i \ \sum j \ W_{ij} \sum i \ (X_i - \bar{X})^2}$$

where: X_i = the SR of cancer for the ith city; $\bar{X}$ = the mean SR of cancer for all of the cities in the study area; X_j = the SR of cancer for the jth city; W_{ij} = a weight parameter for the pair of cities *i* and *j* that represents proximity; and n = the number of cities.

In the study, we apply a simpler version of the same formula of Moran's I for SR as derived by Chen (269).

7.8.2 Steps in computing Moran's Index.

Age-standardized incidence rates of different types of cancers in males and females were retrieved from CI-5 volume X, and the SR of 28 different cancers were computed from cancer registries in Asia and Europe. For each registration area, latitude and longitude coordinates were obtained from a publicly available world geolocation database (270). These geo-coordinates correspond with the approximate center of the geographic area of a cancer registry. These coordinates were used to calculate the geographic *distance in kilometers*

between each pair of registries. In this analysis, we considered the 102 and 52 cancer registries in Europe and Asia respectively. The computations of spatial analysis in the study were carried out using combination of SPSS (version 23.0), R software, and Excel spread sheets.

The flowchart of the spatial autocorrelation analysis in Figure 7-4 shows the steps taken in the calculation of Moran's I. Moran's I was computed for three variables (for each of the 20 cancers): age-standardized incidence rates for males and females, and SR. These three variables are termed as *spatial size measurements (x_1, x_2, and x_3)*, and were calculated for each registry area, separately for Europe and Asia. The means and standard deviations of x_1, x_2, *and* x_3 were calculated to derive standardized variables z_1, z_2, and z_3. The next step involved using latitudes and longitudes to calculate the distance in kilometers to generate spatial distance matrix (Table 7-2). Details of the computations of distances are provided in Appendix 3.

Figure 7-4: A flowchart of steps in computation of Moran's I: data preparation, parameter estimation, and spatial autocorrelation analysis.

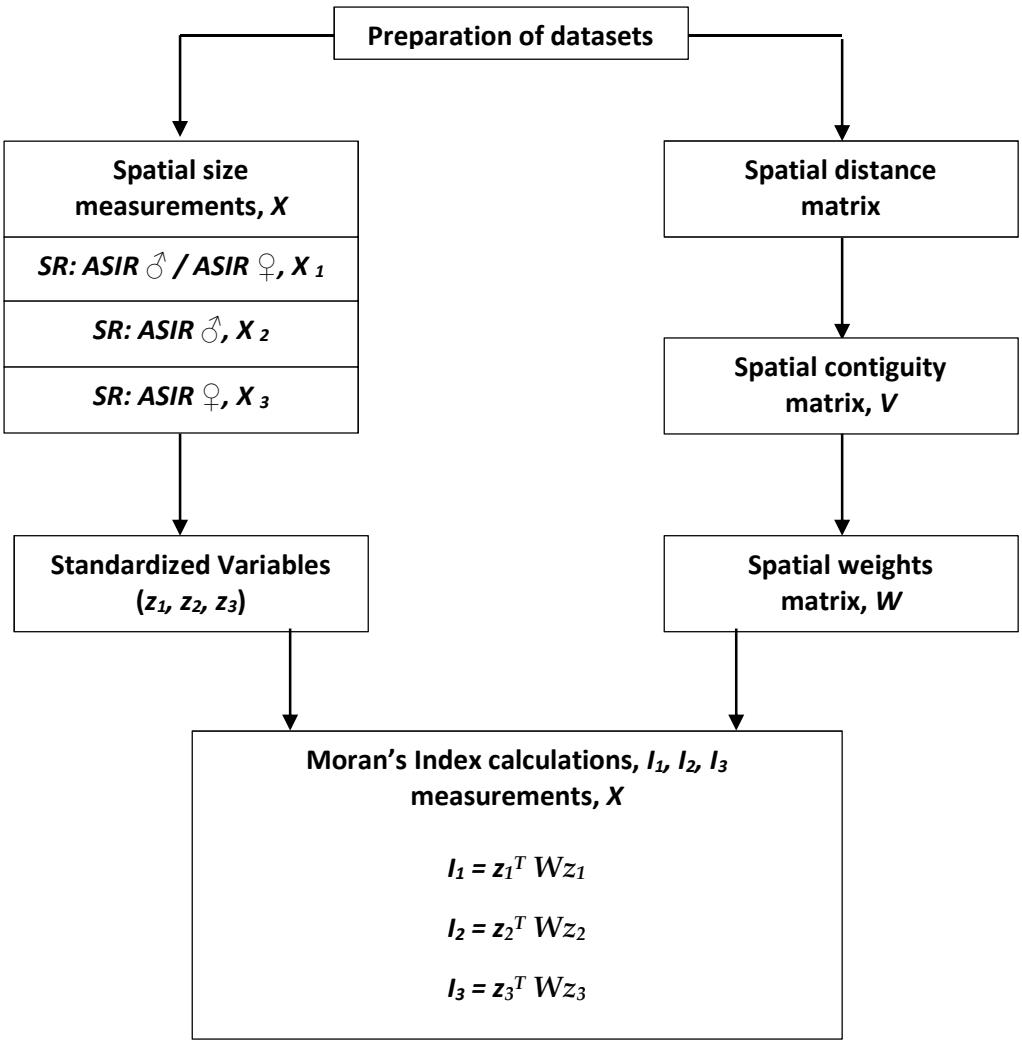


Table 7-2: An example of pair wise spatial distance matrix in kilometers from 10 selected areas in Asia.

(A total of 52 areas in Asia and 102 areas in Europe were used for computing Moran's I in the study)

	Beijing	***Cixian***	***Haining***	***Hong Kong***	***Jiashan***	***Jiaxing***	***Macao***	***Ongang***	***Qidong***	***Shanghai***
Beijing	0	434.99	1124.59	1975.58	186.67	1095.20	1420.67	1891.71	1020.40	1087.68
Cixian	434.99	0	874.68	1566.58	598.43	858.94	1067.75	1477.00	839.85	878.32
Haining	1124.59	874.68	0	1105.57	1172.35	43.52	1572.46	1068.96	189.19	109.04
Hong Kong	1975.58	1566.58	1105.57	0	2095.39	1148.65	1460.37	110.99	1294.61	1212.24
Jiashan	186.67	598.43	1172.35	2095.39	0	1138.20	1606.94	2016.95	1043.02	1119.27
Jiaxing	1095.20	858.94	43.52	1148.65	1138.20	0	1586.44	1111.26	145.96	70.12
Macao	1420.67	1067.75	1572.46	1460.37	1606.94	1586.44	0	1350.15	1657.65	1643.07
Ongang	1891.71	1477.00	1068.96	110.99	2016.95	1111.26	1350.15	0	1256.55	1177.27
Qidong	1020.40	839.85	189.19	1294.61	1043.02	145.96	1657.65	1256.55	0	88.59
Shanghai	1087.68	878.32	109.04	1212.24	1119.27	70.12	1643.07	1177.27	88.59	0

For details of computations of pairwise distances, see Appendix E.

Spatial distance matrix is used for conversion into spatial contiguity matrix *V* and spatial weight matrix *W*

As shown in Table 7-2, distances between any two registry areas are used to create spatial distance matrix, which is then converted into a spatial contiguity matrix V. The spatial contiguity matrix V is created by using a weight function. The type of spatial weight function that is selected is the *inverse power function*, as used by Chen (269, 271):

$$v_{ij} = \begin{cases} r_{ij}^{-1} & ,i \neq j \\ 0, & ,i = j \end{cases} \quad ,$$

where r_{ij} refers to the distance between registry area i and registry area j. Therefore for n registry area, a spatial contiguity matrix, V, can be expressed as:

$$V = [v_{ij}]_{n \times n} = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n} & v_{n2} & \cdots & v_{nn} \end{bmatrix},$$

where v_{ij} measures used to compare and judge the degree of nearness or the contiguous relationships between registry areas i and j (i, j=1,2,…,n). No matter what the entry v_{ii} equals, it will be converted into 0 (for $i=j$, $v_{ii} = 0$). The next step is summation of the spatial contiguity matrix denoted by S using the following formula:

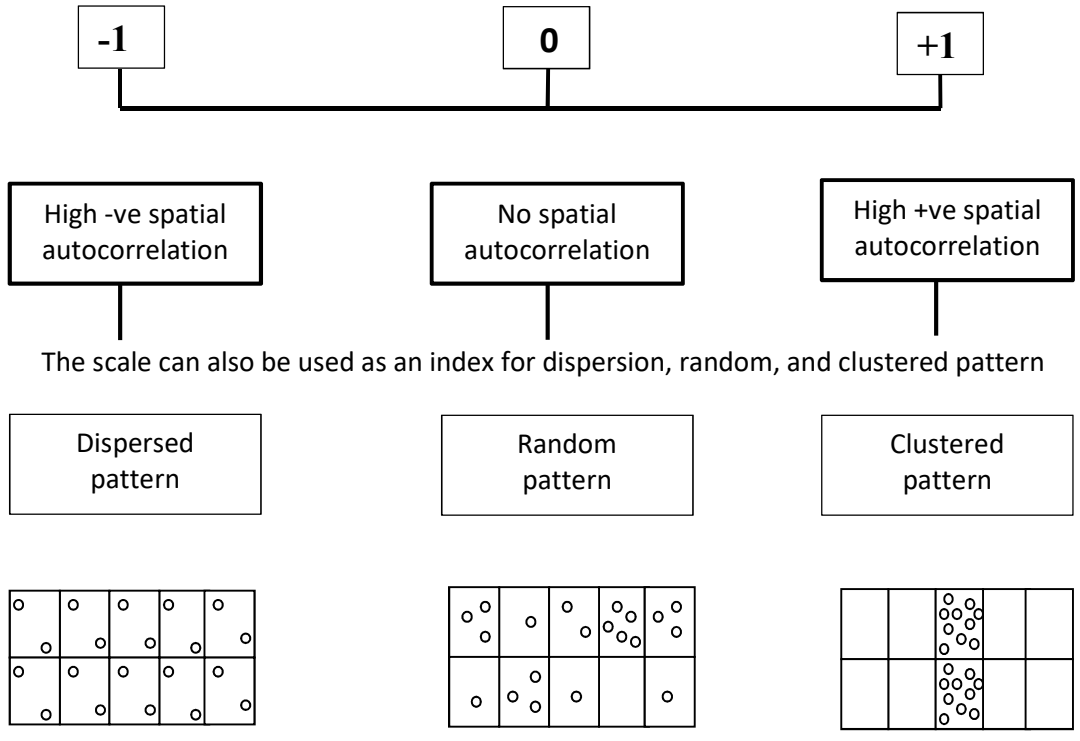
$$S = \sum_{i} \sum_{j} v_{ij} .$$

Therefore, the spatial weights matrix, W, can now be given by the following matrix equation:

$$W = \frac{V}{V_0} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_n & w_{n2} & \cdots & w_{nn} \end{bmatrix}, \qquad \text{where } V_0 = S = \sum_i \sum_j v_{ij}$$

According to the simplified formula provided by Chen (269), $I = z^T Wz$, simplifies the calculation of Moran's I for SRs of cancer incidence as well as for the ASIRs for males and females (I_1, I_2, *and* I_3 as shown in Figure 7-4). The interpretation of Moran's I is carried out on a scale between -1 to +1 as shown in flowchart diagram below:

Scale of Moran's Index:



Source: pro.arc.gis.com ; http://www.utdallas.edu/~briggs/

Chapter 8 Results

The study provides a detailed description of international patterns of cancer occurrence through variability in SR of age-standardized incidence rates worldwide, and then facilitates in conjecturin on the potential etiologic influences. This exercise further demonstrates the utility of SR in the presence of existing artifacts e.g., unequal access to health care services among males and females (i.e., gender-bias). In order to allude onto the potential of gender-bias that could arguably be existing in the cancer registration processes, and that this bias could be operating in different countries where registries are located, comparison of gender inequality and smoking prevalence with SR are also presented for certain cancer types.

Chapter 8 is divided into five sections. Section 8.1 provides descriptive statistics on incidence and the SR of incidence for different cancer types in various cancer registries and time periods. Section 8.2 shows the results of computations of summary measures of SR across registries and time periods, and categories of different types of cancer according to how they manifest in terms of the summary SRm and the SRv. Section 8.3 shows results of mixed effects regression and estimates of time trends and inter- and intra- regional variability of SR. In section 8.4, we also present results of the models incorporating gender inequality and its impact on the sex ratios of selected types of cancers such as cancers of lung, bladder, esophagus, and larynx from 1983 to 2007. This section documents the possibility of gender-bias (or absence thereof) in the study by comparing cancer registries where this bias could be presumably high or low (according to the ranking of gender inequality index). Finally, in Section 8.5, we present results of the spatial analysis of SR, where the cancer types are ranked according to the degree of clustering in Europe and Asia with the observed similarities and dissimilarities among those types. In all the following sections, in some

places, comments are also provided briefly on the observations that were made on the incidence and sex ratio trends across time. We further revisit these observations in Chapter 9.

8.1 Incidence rates and SR worldwide

Incidence rates

Table 8-1 shows cancer types in three different time periods listed from highest to lowest incidence rates in 1973-77 for both sexes combined. The number of registries worldwide varied because of missing incidence data for some cancer types. The registries reporting incidence rates increased from a maximum of 99 in 1973-77 to 142 and 281 in the period 1988-93 and 2003-07, respectively. The results in this Table serve as background data to describe SR of cancer incidence.

Table 8-1 shows that among the top ten cancers with the highest incidence in 1973-77, cancers of *lung* and *stomach* show downward trends (from $32.80/10^5$ in 1973-77 to $30.47/10^5$ in 2003-07 in *lung*; and from 17.56 to $11.35/10^5$ in *stomach*). This may be the result of the declining prevalence of smoking for *lung* cancers and *helicobacter* infections for cancers of *stomach* in much of the developed world (272, 273). Cancers of *colon* as well as *rectum/anus* remained stable throughout the three periods, which can be attributed to stable patterns of risk factors (274). Some of these observations and speculation thereof are discussed in Chapter 9.

The incidence rate of *pancreatic cancer* has a fluctuating pattern. A slightly increasing incidence rates are observed for *leukemia* (6.12 to 7.41 per 10^5). A noticeable increase of 4.82 to 8.95 per 10^5 was observed for *Non-Hodgkin lymphoma (NHL)*. Of all cancers, *bladder* cancer showed the most pronounced increase of 11.19 per 10^5 in 2003-07 from 4.48

per 10^5 in 1973-77. More recently, *liver* cancer shows a high incidence rate of 7.16 per 10^5. A similar pronounced increase such as for *bladder* cancer was observed for *skin melanoma* (8.30 per 10^5) and *thyroid* cancers (6.52 per 10^5) in 2003-07 compared to 1973-77. The *Gallbladder* cancer rate is stable over the three time periods, with only a slight decrease in 2003-07.

Most rare types of cancers with low incidence rates are stable throughout the time periods (e.g., *multiple myeloma, lip, mouth, connective tissue, tongue, bone, salivary gland, nose and sinuses, eye*, and *small intestine)*. The age-standardized incidence rates are graphically depicted in males and females for cancers with highest (Figure 8-1), medium (Figure 8-2), and lowest (Figure 8-3) incidence rates in 1973-77, and changes in those rates thereafter in 1988-92 and 2003-07.

Table 8-1: Total number of cancer registries N and incidence rates *I* in periods 1973-77, 1988-93, and 2003-07 for 30 types of cancers for both sexes.

Cancers*	1973-77		1988-93		2003-07	
	N	$I / 10^5$	N	$I / 10^5$	N	$I / 10^5$
Lung	77	32.80	142	32.66	281	30.47
Skin (non-melanoma)	42	21.74	119	26.41	268	21.61
Stomach	77	17.56	142	14.69	281	11.35
Colon	77	16.00	142	15.71	281	17.53
Rectum & anus	77	9.91	142	9.99	281	9.79
Kidney	74	8.39	141	5.82	281	6.59
Pancreas	77	6.23	142	5.50	281	5.89
Leukemia	77	6.12	142	6.36	254	7.41
Non-Hodgkin lymp.	77	4.82	142	7.26	281	8.95
Bladder	77	4.48	140	10.06	281	11.19
Brain	77	4.45	142	4.62	280	4.83
Esophagus	74	4.41	141	4.27	280	5.20
Larynx	72	3.75	137	3.90	278	3.07
Liver	77	3.59	140	6.22	280	7.16
Skin (melanoma)	74	3.34	139	5.96	277	8.30
Pharynx	77	3.02	140	3.62	146	3.12
Thyroid	76	2.74	140	2.80	281	6.52
Gallbladder	75	2.69	138	2.69	278	2.53
Multiple myeloma	76	2.23	139	2.36	280	2.69
Lip	75	2.11	107	1.60	224	0.70
Hodgkin lymphoma	75	2.10	137	1.90	277	2.10
Mouth	54	1.91	141	2.04	278	1.68
Connective tissue	76	1.55	140	1.69	279	1.90
Tongue	70	1.46	140	1.54	280	1.63
Bone	72	0.94	140	0.95	278	1.00
Salivary gland	66	0.79	133	0.56	273	0.63
Nose & Sinuses	72	0.71	134	0.58	271	0.48
Eye	66	0.62	131	0.62	266	0.62
Small intestine	70	0.57	133	0.61	272	0.88
Endocrine glands	65	0.45	124	0.49	174	0.21

*Cancer types are sorted by reported incidence from highest to lowest in 1973-77

Figure 8-1 shows the ten cancers with the highest incidence in 1973-77, each of which (with the exceptions of *lung* and *bladder cancer)* had a similar pattern over time in males and females. The incidence rate of *lung* cancer decreased over time in males (52.74 to 43.04 / 10^5); it increased in females (12.86 to 17.91/10^5). Differences in *lung* cancer incidence patterns across sex could reflect historical differences in tobacco use (272, 273). Women took up smoking in large numbers later than men, initiated smoking at older ages, and were slower to quit, including recent upturns in smoking prevalence in some birth cohorts (273, 275). Declines in lung cancer incidence continue to be larger in men than in women.

Figure 8-2 shows 10 cancers with mid-level incidence rates. *Thyroid* cancer was the most rapidly increasing cancer in women, partially due to over-diagnosis because of increased use of advanced imaging techniques. Increases in tumor size and stage, in follicular carcinoma (a more aggressive subtype), suggest that some of the rise may be due to changes in environmental risk factors such as obesity (273, 276, 277). These observations with debatable speculations form part of discussion in Chapter 9. Figure 8-3 shows the 10 cancers with the lowest incidence, where the patterns are similar across sex.

Most ecological analyses and international studies of cancer incidence time-trends have provided somewhat similar results previously (273, 275-277). We believe that these trends, as informative as they are, should be interpreted with caution as there are reasons to believe that any comparisons worldwide are likely to be confounded by many factors such as addition of new population-based registries, increases in older population or increased access to diagnostic services. This suspicion has prompted us to conduct a detailed analysis of the SR in incidence rates, although the issue of different number of registries in three-time periods still remains.

Figure 8-1: Cancers with the highest incidence rates in 1973-77 (> 4.5 / 10^5 person-years) and changes in incidence rates over time in 1988-92 and 2003-07, by sex.

(Males represented by solid line and females by dashed line).

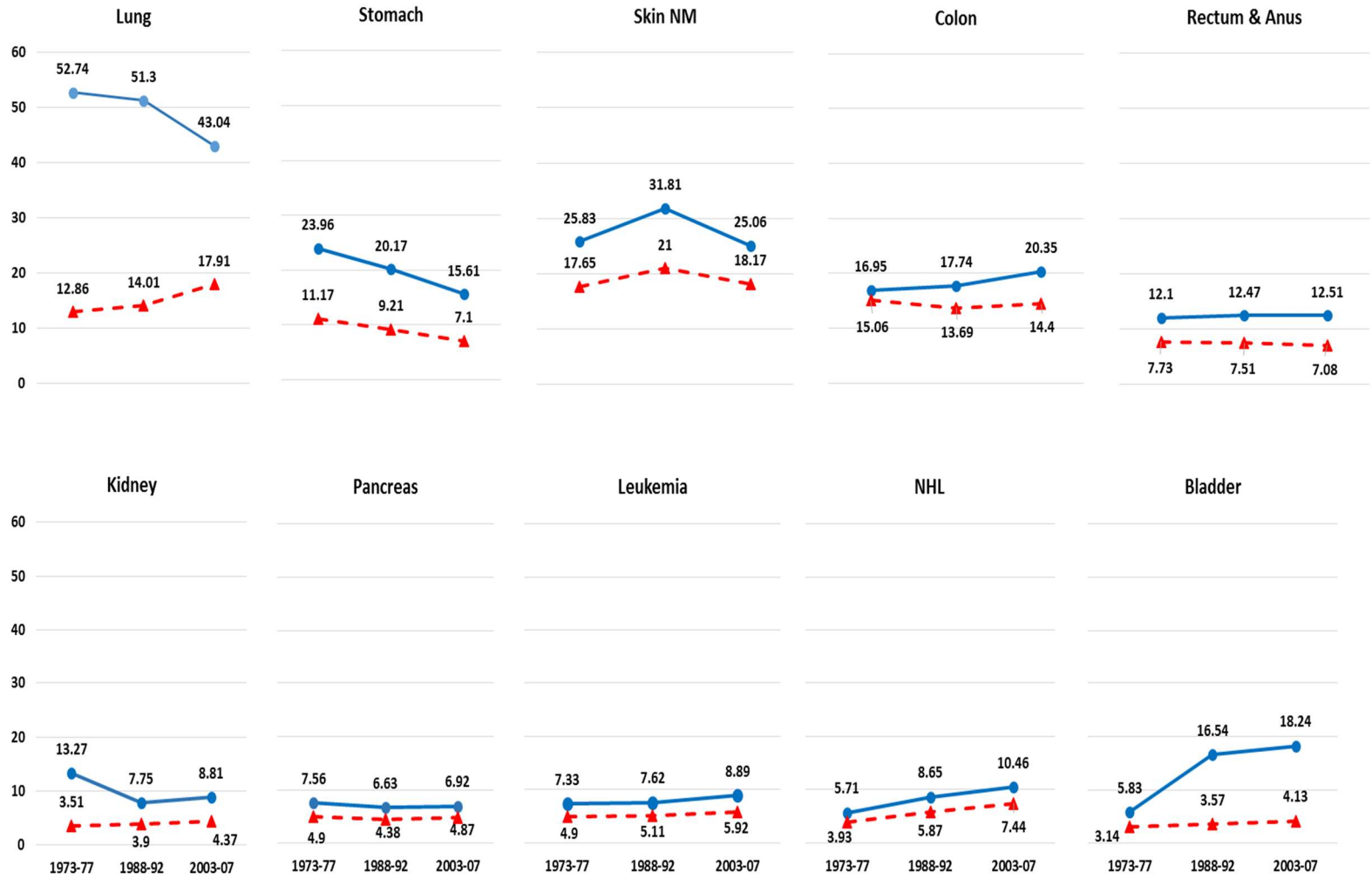


Figure 8-2: Cancers with medium-level incidence rates in 1973-77 (2.0 - 4.5 / 10^5 person-years) and changes in incidence rates over time in 1988-92 and 2003-07, by sex

(Males represented by solid line and females by dashed line).

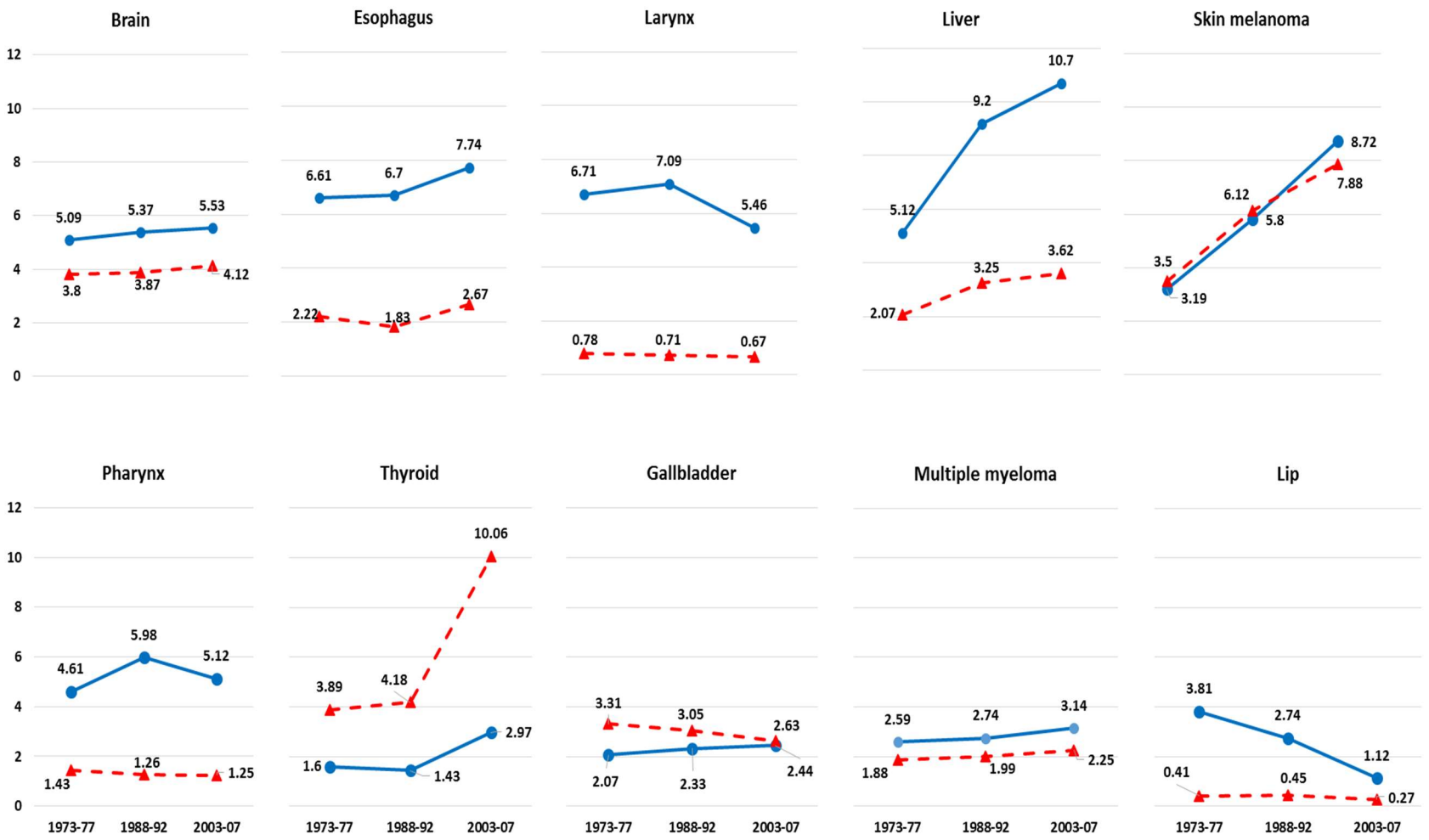
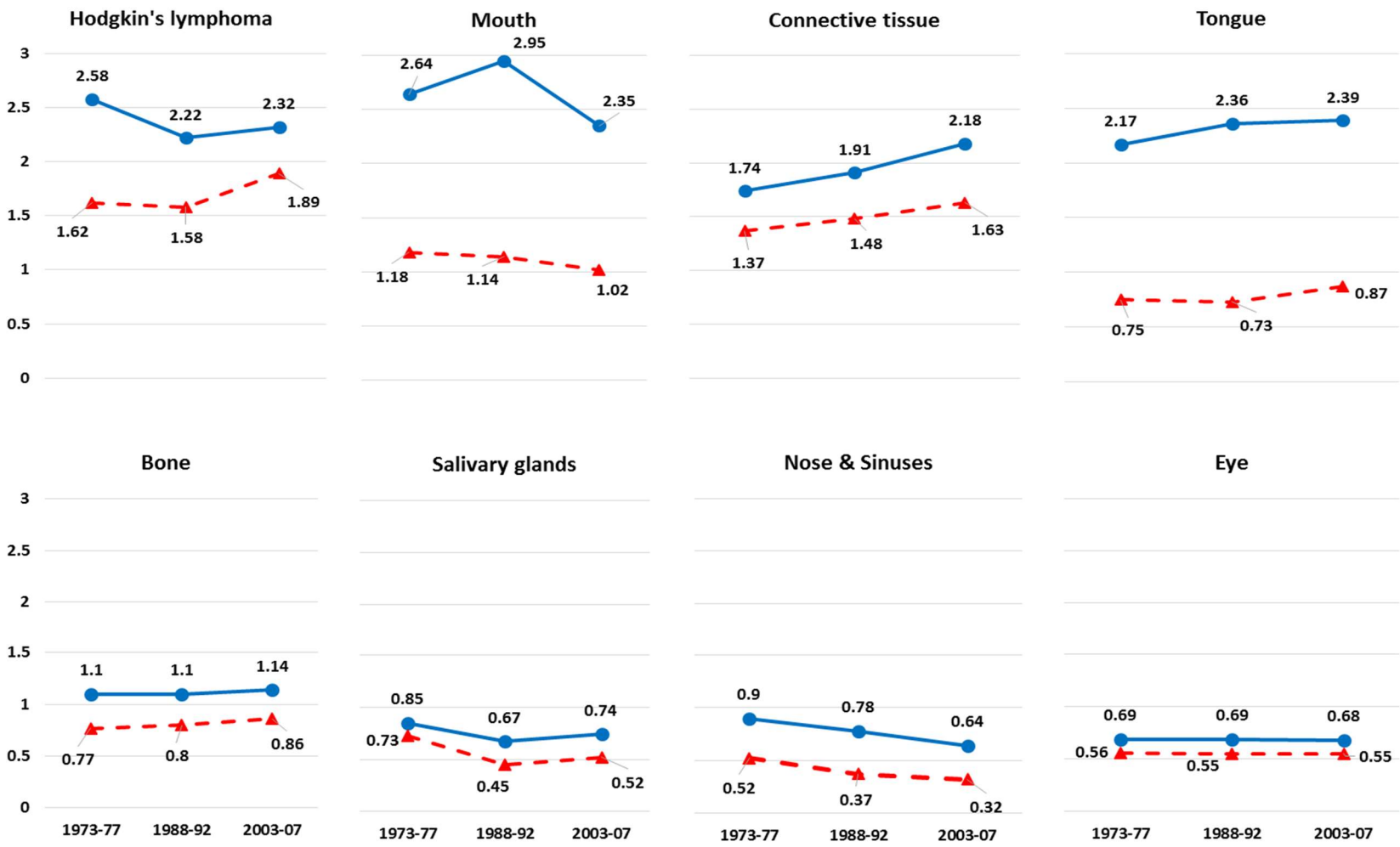


Figure 8-3: Cancers with low incidence rates in 1973-77 (< 3 / 10^5 person-years) and changes in incidence rates over time in 1988-92 and 2003-07, by sex

(Males represented by solid line and females by dashed line).



SR of incidence rates

The ratio of ASIR in males relative to females (termed herein as the magnitude of SR (SRm), in three time periods for different types of cancers are shown in Table 8-2. The detailed computation on SRm denoted by SRm_i^k is already provided in Chapter 7 (Table 7-1).

SRm is indicative of the predominance of one sex over the other for a specific cancer type. The highest overall magnitude over three time-periods was observed for cancer of the *larynx* (13.9, 15.3, and 11.5) The SR suggested that it is a highly male dominant cancer. The lowest magnitudes in 1973-77, 1988-89, and 2003-07 were noted for cancers of the *thyroid* (0.4, 0.4, and 0.3) and *gallbladder* (0.8, 0.8, and 0.9) showing female dominance.

A changing or steady pattern of SRm suggests an underlying pattern of known or unknown exposures in different geographic regions, which is why we report, in the latter section, variations in SRm or SRv. Other notable trends in SRm were observed in cancers of the *lung* and *kidney, which* indicate a rise in female incidence rates (e.g., SRm was low in 2003-07 for *lung* cancer (SRm = 3.3) compared to 5.3 in 1973-77). The rise in female kidney cancer however mirrored the increase in male incidence as seen in Figure 8-3, hence it remains stable at a sex ratio of 2 (Table 8-2). For *bladder* cancer, SRm increased in the last two time periods (5.1 and 4.7) from 2.0 in 1973-77. A fluctuating SRm was observed for *pharyngeal* cancer with a peak of 6.1 in 1988-93 compared to 4.3 in 1973-77 and 4.8 in 2003-07. Most other low-incident cancers, were fairly stable over time. To better illustrate variance in SRm across registries, Tables 8-3 to 8-5 show the registries with the highest and lowest SRm for the same 30 cancers.

Table 8-2: The sex ratio magnitude (SRm) of 30 types of cancers worldwide in 1973-77, 1988-93, and 2003-07.

Cancer	SRm in three time periods		
	1973-77	1988-93	2003-07
Lung	5.3	5.4	3.3
Skin (non-melanoma)	1.5	1.6	1.4
Stomach	2.2	2.2	2.2
Colon	1.1	1.3	1.4
Rectum & anus	1.6	1.7	1.7
Kidney	4.2	2.0	2.1
Pancreas	1.6	1.6	1.4
Leukemia	1.5	1.5	1.5
Non-Hodgkin lymphoma	1.6	1.5	1.5
Bladder	2.0	5.1	4.7
Brain	1.5	1.5	1.4
Esophagus	4.1	5.4	5.0
Larynx	13.9	15.3	11.5
Liver	2.5	3.0	3.0
Skin (melanoma)	1.1	1.0	1.1
Pharynx	4.3	6.1	4.8
Thyroid	0.4	0.4	0.3
Gallbladder	0.8	0.8	0.9
Multiple myeloma	1.5	1.4	1.4
Lip	10.7	7.4	4.6
Hodgkin lymphoma	1.7	1.5	1.4
Mouth	3.3	3.7	2.5
Connective tissue	1.4	1.4	1.4
Tongue	3.6	4.2	3.0
Bone	1.6	1.5	1.4
Salivary gland	1.3	1.8	1.6
Nose & Sinuses	2.0	2.6	2.3
Eye	1.4	1.4	1.5
Small intestine	1.5	1.7	1.6
Endocrine glands	1.6	1.3	1.7

*Cancer types are sorted by reported incidence rates from highest to lowest in 1973-77

Table 8-3: Incidence rates of 30 cancers in males and females, the sex ratio magnitude (SRm), and registries with the highest and lowest SRm in 1973-77.

Cancer	Incidence		SRm	Highest SRm		Lowest SRm	
	Male	Female	*(Overall)*				
Lung	52.74	12.86	5.3	France	15.1	Canada	2.1
Skin (non-melanoma)	25.83	17.65	1.5	Spain	2.3	India	0.9
Stomach	23.96	11.17	2.2	Switzerland	4.5	India	1.1
Colon	16.95	15.06	1.1	Romania	1.6	Canada	0.7
Rectum & anus	12.10	7.73	1.6	Canada	3.0	Cuba	1.1
Kidney	13.27	3.51	4.2	France	11.6	Canada	0.7
Pancreas	7.56	4.90	1.6	Netherland	3.6	Canada	0.3
Leukemia	7.33	4.90	1.5	Senegal	3.5	Canada	0.6
Non-Hodgkin lymphoma	5.71	3.93	1.6	Poland	3.1	Switzerland	0.8
Bladder	5.83	3.14	2.0	Japan	4.3	Canada	0.5
Brain	5.09	3.80	1.5	Switzerland	2.3	Canada	0.9
Esophagus	6.61	2.22	4.1	France	21.3	Senegal	1.0
Larynx	6.71	0.78	13.9	France	112.0	Scotland	3.5
Liver	5.12	2.07	2.5	Switzerland	7.5	Canada	0.4
Skin (melanoma)	3.19	3.50	1.1	Spain	2.0	Scotland	0.4
Pharynx	4.61	1.43	4.3	France	31.0	Poland	0.9
Thyroid	1.60	3.89	0.4	Poland	1.1	Poland	0.1
Gallbladder	2.07	3.31	0.8	Scotland	1.4	Hungary	0.2
Multiple myeloma	2.59	1.88	1.5	China	6.0	Japan	0.6
Lip	3.81	0.41	10.7	Spain	59.0	Senegal	0.3
Hodgkin lymphoma	2.58	1.62	1.7	Australia	3.5	France	0.3
Mouth	2.64	1.18	3.3	Slovenia	15.5	Senegal	0.8
Connective tissue	1.74	1.37	1.4	Canada	5.7	Poland	0.4
Tongue	2.17	0.75	3.6	Slovenia	30.0	China	0.8
Bone	1.10	0.77	1.6	Netherlands	8.0	Senegal	0.7
Salivary gland	0.85	0.73	1.3	Spain	3.0	Canada	o.4
Nose & Sinuses	0.90	0.52	2.0	France	6.6	Senegal	1.0
Eye	0.69	0.56	1.4	Canada	3.5	Scotland	0.4
Small intestine	0.65	0.49	1.5	Italy	4.0	Colombia	0.5
Endocrine glands	0.48	0.41	1.6	Netherlands	5.0	Switzerland	0.3

*Cancer types are sorted by reported incidence rates from highest to lowest in 1973-77.

Table 8-4: Incidence rates of 30 cancers in males and females, the sex ratio magnitude (SRm), and registries with the highest and lowest SRm in 1988-92.

Cancer	Incidence		SRm	Highest SRm		Lowest SRm	
	Male	Female	*(Overall)*				
Lung	51.30	14.01	5.4	Spain	17.8	Thailand	1.2
Skin (non-melanoma)	31.81	21.00	1.6	Korea	5.0	Spain	0.6
Stomach	20.17	9.21	2.2	Algeria	4.1	Zimbabwe	0.8
Colon	17.74	13.69	1.3	India	2.4	Canada	0.6
Rectum & anus	12.47	7.51	1.7	India	5.3	Brazil	0.6
Kidney	7.75	3.90	2.0	India	4.0	Uganda	0.5
Pancreas	6.63	4.38	1.6	India	10.5	Zimbabwe	0.7
Leukemia	7.62	5.11	1.5	Korea	3.1	Mali	0.3
Non-Hodgkin lymphoma	8.65	5.87	1.5	India	4.6	Kuwait	0.8
Bladder	16.54	3.57	5.1	Algeria	18.0	Zimbabwe	1.1
Brain	5.37	3.87	1.5	Canada	10.6	French Polynesia	0.7
Esophagus	6.70	1.83	5.4	Spain	31.0	Canada	0.6
Larynx	7.09	0.71	15.3	Spain	171.0	Uganda	0.4
Liver	9.20	3.25	3.0	France	10.1	Peru	1.0
Skin (melanoma)	5.80	6.12	1.0	Japan	3.0	Mali	0.3
Pharynx	5.98	1.26	6.1	Spain	25.0	Canada	0.8
Thyroid	1.43	4.18	0.4	India	2.5	Spain	0.1
Gallbladder	2.33	3.05	0.8	Vietnam	3.0	Brazil	0.3
Multiple myeloma	2.74	1.99	1.4	India	6.5	India	0.5
Lip	2.74	0.45	7.4	Canada	38.0	Thailand	0.1
Hodgkin lymphoma	2.22	1.58	1.5	Peru	9.0	Japan	0.5
Mouth	2.95	1.14	3.7	Switzerland	22.0	India	0.3
Connective tissue	1.91	1.48	1.4	Canada	9.0	Brazil	0.5
Tongue	2.36	0.73	4.2	Slovakia	17.0	China	0.5
Bone	1.10	0.80	1.5	Switzerland	6.5	Spain	0.3
Salivary gland	0.67	0.45	1.8	France	7.0	Malta	0.2
Nose & Sinuses	0.78	0.37	2.6	Italy	10.0	Italy	0.5
Eye	0.69	0.55	1.4	Italy	5.0	India	0.5
Small intestine	0.72	0.50	1.7	Spain	10.0	Zimbabwe	0.3
Endocrine glands	0.51	0.46	1.3	Estonia	4.0	Iceland	0.3

*Cancer types are sorted by reported incidence rates from highest to lowest in 1973-77

Table 8-5: Incidence rates of 30 cancers in males and females, the sex ratio magnitude (SRm), and registries with the highest and lowest SRm in 2003-07.

Cancer	Incidence		SRm	Highest SRm		Lowest SRm	
	Male	Female	*(Overall)*				
Lung	43.04	17.91	3.3	Turkey	15.6	India	0.7
Skin (non-melanoma)	25.06	18.17	1.4	France	4.0	Spain	0.2
Stomach	15.61	7.10	2.2	India	4.8	Canada	0.7
Colon	20.35	14.40	1.4	India	3.7	India	0.7
Rectum & anus	12.51	7.08	1.7	Argentina	3.8	Ecuador	0.6
Kidney	8.81	4.37	2.1	India	8.0	China	0.5
Pancreas	6.92	4.87	1.4	Iran	2.8	Malawi	0.3
Leukemia	8.89	5.92	1.5	China	2.5	Portugal	0.9
Non-Hodgkin lymphoma	10.46	7.44	1.5	China	4.1	India	0.5
Bladder	18.24	4.13	4.7	India	15.0	Canada	1.0
Brain	5.53	4.12	1.4	China	2.2	Malawi	0.3
Esophagus	7.74	2.67	5.0	Spain	25.0	Qatar	0.6
Larynx	5.46	0.67	11.5	Spain	100.0	Canada	0.5
Liver	10.70	3.62	3.0	Argentina	12.2	Ecuador	0.7
Skin (melanoma)	8.72	7.88	1.1	India	4.0	China	0.3
Pharynx	5.12	1.25	4.8	Spain	47.0	China	1.1
Thyroid	2.97	10.06	0.3	Canada	1.0	Argentina	0.0
Gallbladder	2.44	2.63	0.9	France	2.4	India	0.3
Multiple myeloma	3.14	2.25	1.4	Canada	7.0	India	0.3
Lip	1.12	0.27	4.6	Australia	23.0	Thailand	0.2
Hodgkin lymphoma	2.32	1.89	1.4	India	9.0	China	0.3
Mouth	2.35	1.02	2.5	Belarus	12.3	China	0.3
Connective tissue	2.18	1.63	1.4	China	4.5	China	0.3
Tongue	2.39	0.87	3.0	Belarus	15.0	Colombia	0.3
Bone	1.14	0.86	1.4	Italy	5.3	Turkey	0.3
Salivary gland	0.74	0.52	1.6	Italy	12.0	Canada	0.2
Nose & Sinuses	0.64	0.32	2.3	Spain	10.0	Uganda	0.2
Eye	0.68	0.55	1.5	Switzerland	14.0	Korea	0.3
Small intestine	1.03	0.72	1.6	Korea	9.0	Argentina	0.3
Endocrine glands	0.24	0.18	1.7	Italy	7.0	Spain	0.3

*Cancer types are sorted by reported incidence rates from highest to lowest in 1973-77

In 1973-77, among the total number of cancer registries (range 65 to 99), a very large number of them belonged to European regions (Table 8-3). The highest SRm were observed for the following cancers: *larynx* (112.0 in France) followed by *lip* (59.0 in Spain), *pharynx* (31.0 in France), *tongue* (30.0 in Slovenia), *esophagus* (21.3 in France), *mouth* (15.5 in Slovenia), *lung* (15.1 in France), and *kidney* (11.6 in France). The lowest SRm was noted in the cancer registry in Poland for *thyroid* cancer (0.1). In 1973-77, cancer of the *liver* had both the highest and lowest SRm in different cancer registries in Canada (7.5 and 0.4). Table 8-4 shows the highest and lowest SRm from cancer registries in 1988-93. The highest SRm for this period were observed for the following cancers: *larynx* (171.0 in Spain) followed by *lip* (38.0 in Canada), *esophagus* (31.0 in Spain), *pharynx* (25.0 in Spain), *mouth* (22.0 in Switzerland), *bladder* (18.0 in Algeria), *lung* (17.8 in Spain), and *tongue* (17.0 in Slovakia). The lowest SRm was noted from a cancer registry in Spain for *thyroid* cancer (0.1) and Thailand for *lip* (0.1). Other important cancers from the public health burden perspective included *liver* cancer, which showed the highest SRm in France (10.1) and lowest in Peru (1.0). Algeria had the highest SRm for stomach cancer (4.1) whereas the lowest was observed in Zimbabwe (0.8). Zimbabwe also showed the lowest SRm of 0.7 for *pancreatic* cancer, and India had the highest SRm of 10.5. During 1988-92, Canada had the highest SRm for *brain* cancer and the lowest value was observed in French Polynesia (0.7).

A list of registries with the highest and lowest SRm in 2003-07 (a time period when the number of cancer registries representing different countries increased to 281) is presented in Table 8-5. The highest SRm were observed for following cancers: *larynx* (100 in Spain) followed by *pharynx* (47.0 in Spain), *esophagus* (20.0 in Spain), *lip* (23.0 in Australia), *lung* (15.6 in Turkey), *tongue* (15.0 in Belarus), and *bladder* (15.0 in India). *Thyroid* cancer had

the lowest SRm in Argentina whereas registries in Spain, Canada, and Uganda showed some of the lowest SRm for *non-melanoma skin*, *salivary glands*, and *nose and sinuses*. For *liver* cancers, the highest SRm was seen in Argentina (12.2) and the lowest in Ecuador (0.7). Compared to 1973-77, SRm of *Hodgkin's lymphoma* was very high in both 1988-92 and 2003-07 (9.0 in Peru and India). Another notable finding in 2003-07 is a very high SRm for cancer of the eye (14.0 in Switzerland) compared to 1988-93 (5.0 in Italy) and 1973-77 (3.5 in Canada).

Appendix 4 provides a detailed list of cancer registries representing different regional populations in the three time periods, with the highest and lowest SRm (in the column labelled overall SRm, variation SRv is provided in parenthesis).

Descriptive results on incidence rates and their SR in different time periods in diverse geographic regions provide information on male- or female-predominance across numerous cancer types. The next section presents results for SRm and SRv together (in the form of 3 x 3 matrices) that could provide an understanding and some clues on possible geographical variations in known and potential causes of cancers.

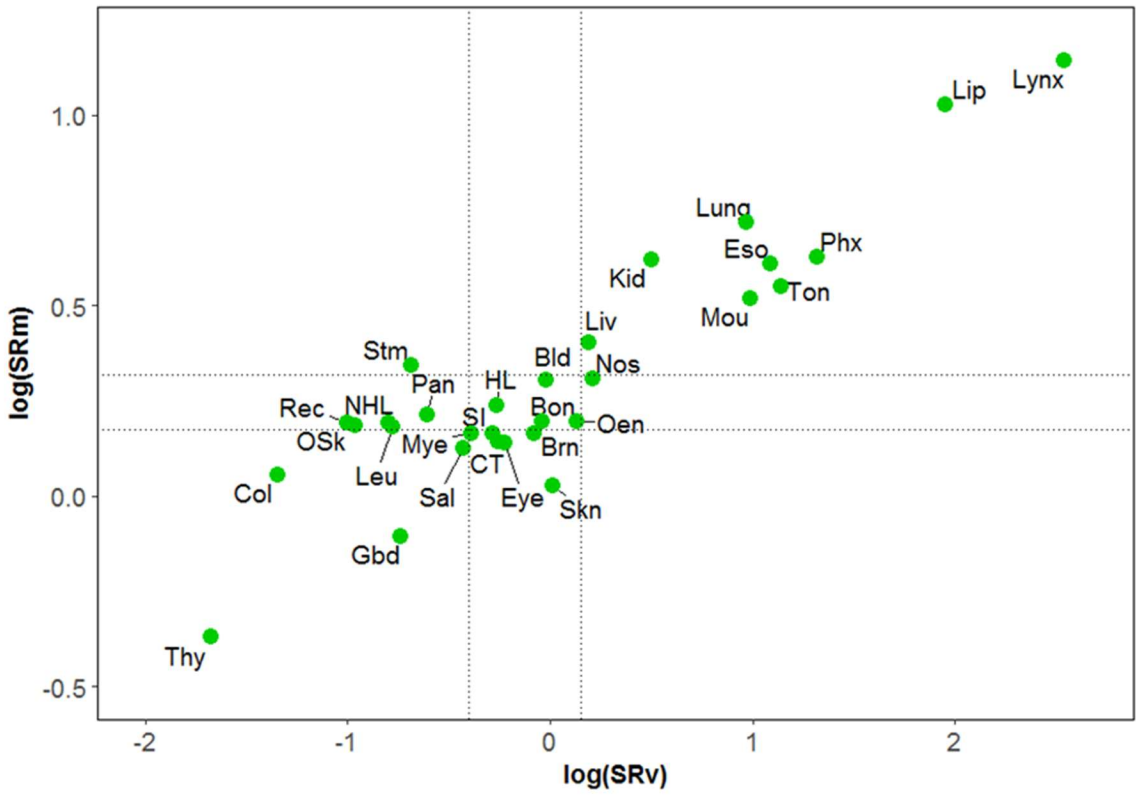
8.2 Cancer types by SRm and SRv.

We believe that the stability or conversely, inconsistency of SR across countries and time periods is an important characteristic of a given disease that may convey information about etiology. For this reason, we computed the international SRv for cancer. However the amplitude of SRv may relate to the SRm and to the incidence of the cancer type in the region. For this reason, we conducted a series of analyses estimating the SRv as well as its magnitude. While incidence, SRm and SRv are computed as continuous variables, we believe

it is useful for the purpose of characterizing cancer types, to categorize each metric. We chose to create 3 categories for each metric as a compromise between preserving some quantification, but also retaining sufficient numbers. The decision to create 3 categories led to the creation of Tables with 9 (3x3) or 27 (3x3x3) categories in which different cancer types are categorized.

The results in this section are based on methods described in Section 7.5. Graphical depictions of the SRm and SRv (on logarithmic scale) in 1973-77 are shown in Figure 8-4 and Table 8-6. Table 8-6 used the templates in Figure 7-2 to present cancers with low, medium and high levels of SRm and SRv in 1973-77. The results are also stratified according to three levels of incidence rates of cancers worldwide using template in Figure 7-3 (Table 8-7). Figures and Tables of SRm and SRv with stratifications by incidence are also shown for 1988-92 (Figure 8-5, Tables 8-8 and 8-9) and for 2003-07 (Figure 8-6, Tables 8-10 and 8-11). The raw version (i.e., untransformed or on arithmetic scale) of these results are presented in Appendix 5. Both transformed and untransformed data highlight similar patterns of cancer types over time (i.e., in 1973-77, 1988-93 and 2003-07).

Figure 8-4: Magnitude of the sex ratios (SRm) of 30 cancer types plotted against their variances (SRv) on log scale for period 1973-1977



Bladder (Bld); Bone (Bon); Brain (Brn); Colon (Col); Connective Tissue (CT); Eye (Eye); Gallbladder (Gbd); Hodgkin Lymphoma (HL); Kidney (Kid); Larynx (Lynx); Leukemia (Leu); Lip (Lip); Liver (Liv); Lung (Lung); Melanoma of Skin (Skn); Mouth (Mou); Multiple Myeloma (Mye); Non Hodgkin Lymphoma (NHL); Nose and Sinuses (Nos); Oesophagus (Eso); Other endocrine cancers (Oen); Other Skin cancers (OSk); Pancreas (Pan); Pharynx (Phx); Rectum and Anus (Rec); Salivary glands (Sal); Small Intestine (SI); Stomach (Stm); Thyroid (Thy); Tongue (Ton).

Table 8-6: Cancer types categorized according to three levels (low, medium & high) of the sex ratio magnitude (SRm) and variances (SRv) in 77 registries in 1973-1977 on log scale.

Log SRm	Log SRv: Low	Log SRv: Medium	Log SRv: High
Low	Colon, Gallbladder, Thyroid, Multiple myeloma	Salivary glands, Small Intestine, skin Melanoma, Connective Tissue, Eye, Brain	
Medium	Rectum & anus, Non-Hodgkin Lymphoma, Leukemia, Other Skin ca.	Bone, Bladder, Hodgkin Lymphoma	Nose & Sinuses, Other endocrine ca.
High	Stomach	Liver	Lip, Tongue, Mouth, Pharynx, Oesophagus, Larynx, Lung, Kidney

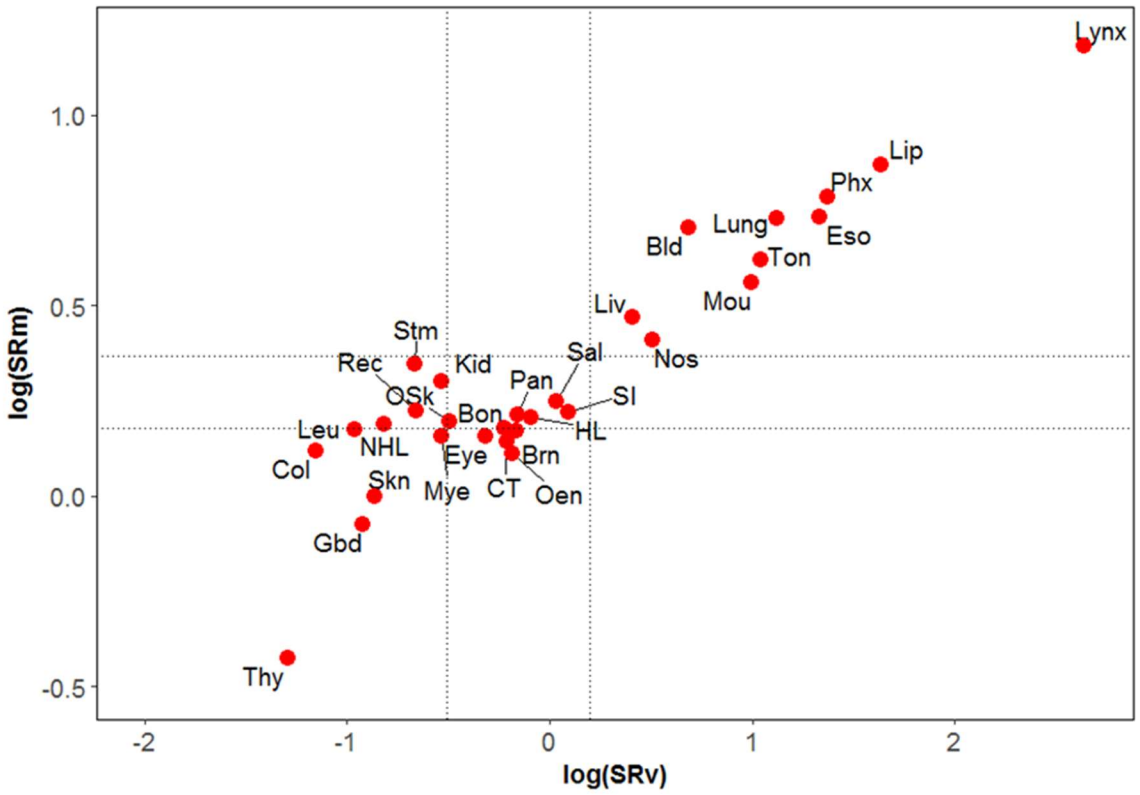
SRm_{ik} (Low): ≤ 0.176; SRm_{ik} (Medium): 0.177 to 0.319; SRm_{ik} (High): > 0 .319
SRv_{ik} (Low): ≤ - 0.405; SRv_{ik} (Medium): - 0.405 to 0.152; SRv_{ik} (High): > 0.152

Table 8-7: Cancer sites stratified according to low, medium and high levels of incidences rates, sex ratio magnitude (SRm) and variances (SRv) in 77 registries in 1973-1977 on log scale.

		Low incidence			**Medium incidence**			**High incidence**		
		Log SRv			***Log SRv***			***Log SRv***		
		Low	***Medium***	***High***	***Low***	***Medium***	***High***	***Low***	***Medium***	***High***
Log SRm	***Low***		Salivary glands Small Intestine, Connective tissue, Eye		Gallbladder, Thyroid, Multiple Myeloma	Skin Melanoma, Brain		Colon		
	Medium		Bone, Hodgkin-Lymphoma,	Nose & Sinuses, endocrine ca.				Rectum & Anus, Non Hodgkin-Lymphoma, Leukemia, Other Skin Ca.	Bladder	
	High			Tongue, Mouth		Liver	Lip, Pharynx, Oesophagus, Larynx	Stomach		Lung, Kidney

Low incidence: $\leq 1.63 / 10^5$; Medium incidence: 1.64 to $6.52 / 10^5$; High incidence: $> 6.52 / 10^5$
SRm_{ik} (Low): ≤ 0.52; SRm_{ik} (Medium): 0.53 to 0.61; SRm_{ik} (High): >0 .61
SRv_{ik} (Low): ≤ 0.003); SRv_{ik} (Medium): 0.004 to 0.014; SRv_{ik} (High): > 0.014

Figure 8-5: Magnitude of sex ratios (SRm) of 30 cancer types plotted against their variances (SRv) on log scale for period 1988-1992.



Bladder (Bld); Bone (Bon); Brain (Brn); Colon (Col); Connective Tissue (CT); Eye (Eye); Gallbladder (Gbd); Hodgkin Lymphoma (HL); Kidney (Kid); Larynx (Lynx); Leukemia (Leu); Lip (Lip); Liver (Liv); Lung (Lung); Melanoma of Skin (Skn); Mouth (Mou); Multiple Myeloma (Mye); Non Hodgkin Lymphoma (NHL); Nose and Sinuses (Nos); Oesophagus (Eso); Other endocrine cancers (Oen); Other Skin cancers (OSk); Pancreas (Pan); Pharynx (Phx); Rectum and Anus (Rec); Salivary glands (Sal); Small Intestine (SI); Stomach (Stm); Thyroid (Thy); Tongue (Ton).

Table 8-8: Cancer types categorized according to three levels (low, medium & high) of sex ratio magnitude (SRm) and variances (SRv) in 142 registries in 1988-1992 on log scale.

		Log SRv		
		Low	***Medium***	***High***
Log SRm	***Low***	Colon, Gallbladder, Skin melanoma, Thyroid, Multiple myeloma,	Bone, Connective tissue, Eye, Brain, endocrine cancers,	
	Medium	Stomach, Rectum & Anus, Kidney, Non Hodgkin-Lymphoma, Leukemia	Salivary glands, Small Intestine, Pancreas, Hodgkin Lymphoma, Other skin cancers	
	High			Lip, Tongue, Pharynx, Oesophagus, Liver, Nose & Sinuses, Mouth, Larynx, Lung, Bladder

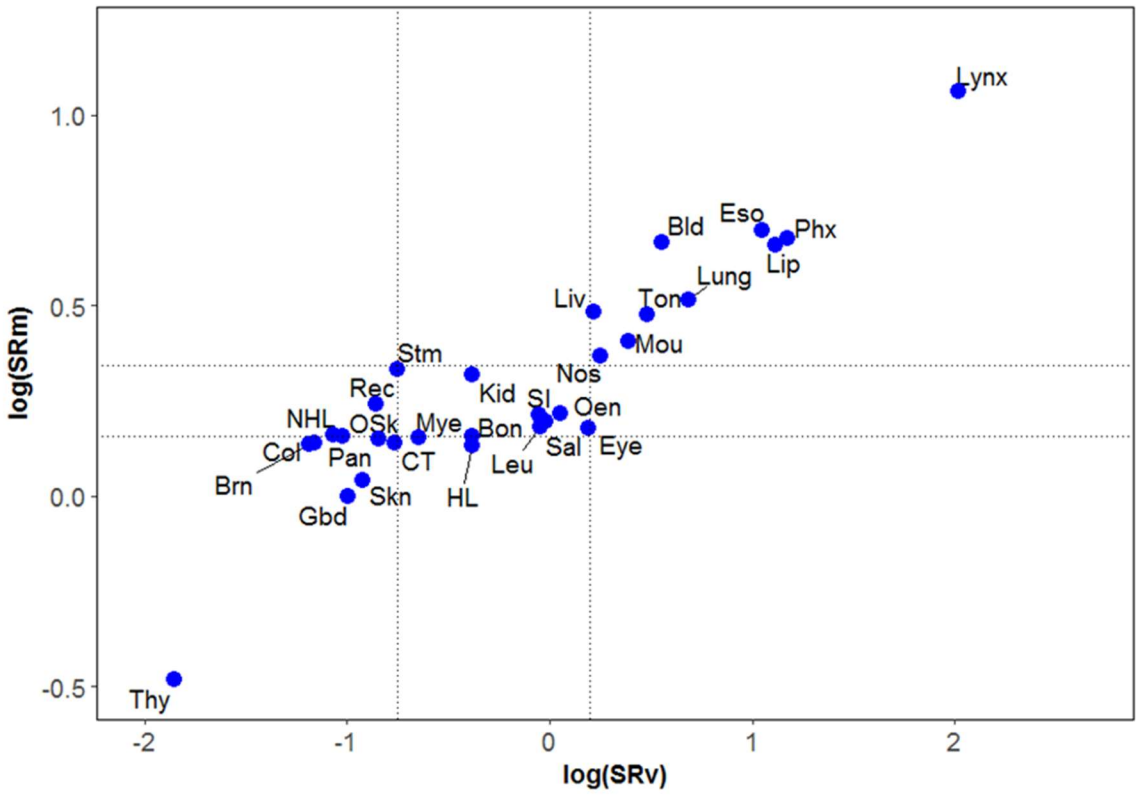
SRm_{ik} (Low): ≤ 0.177; SRm_{ik} (Medium): 0.178 to 0.0.369; SRm_{ik} (High): > 0 .369
SRv_{ik} (Low): ≤ - 0.509; SRv_{ik} (Medium): - 0.509 to 0.198; SRv_{ik} (High): > 0.198

Table 8-9: Cancer sites stratified according to low, medium and high levels of incidence rates, sex ratio magnitude (SRm) and variances (SRv) in 142 registries in 1988-1992 on log scale.

		Low incidence			Medium incidence			High incidence		
		Log SRv			*Log SRv*			*Log SRv*		
		Low	*Medium*	*High*	*Low*	*Medium*	*High*	*Low*	*Medium*	*High*
Log SRm	*Low*		Bone, Connective tissue, Eye, endocrine glands,		Gallbladder, Thyroid, Multiple Myeloma	Brain		Colon, Skin Melanoma,		
	Medium		Salivary glands, Small-Intestine, Hodgkin-Lymphoma,		Kidney	Pancreas		Stomach, Rectum & Anus, Non Hodgkin-Lymphoma, Leukemia	Other skin cancers	
	High			Lip, Tongue, Nose & Sinuses			Mouth, Pharynx, Oesophagus, Larynx			Liver, Lung, Bladder

Low incidence: $\leq 1.90 / 10^5$; Medium incidence: 1.95 to $5.59 / 10^5$; High incidence: $> 5.59 / 10^5$
SRm_{ik} (Low): ≤ 0.53; SRm_{ik} (Medium): 0.54-0.63 to 2.56; SRm_{ik} (High): > 0.63
SRv_{ik} (Low): ≤ 0.003); SRv_{ik} (Medium): 0.004 to 0.0012; SRv_{ik} (High): > 0.012

Figure 8-6: Magnitude of the sex ratios (SRm) of 30 cancer types plotted against their variances (SRv) on log scale for period 2003-07.



Bladder (Bld); Bone (Bon); Brain (Brn); Colon (Col); Connective Tissue (CT); Eye (Eye); Gallbladder (Gbd); Hodgkin Lymphoma (HL); Kidney (Kid); Larynx (Lynx); Leukemia (Leu); Lip (Lip); Liver (Liv); Lung (Lung); Melanoma of Skin (Skn); Mouth (Mou); Multiple Myeloma (Mye); Non Hodgkin Lymphoma (NHL); Nose and Sinuses (Nos); Oesophagus (Eso); Other endocrine cancers (Oen); Other Skin cancers (OSk); Pancreas (Pan); Pharynx (Phx); Rectum and Anus (Rec); Salivary glands (Sal); Small Intestine (SI); Stomach (Stm); Thyroid (Thy); Tongue (Ton).

Table 8-10:Cancer types categorized according to three levels (low, medium & high) of sex ratio magnitude (SRm) and variances (SRv) in 281 registries in 2003-2007 on log scale.

		Log SRv		
		Low	***Medium***	***High***
Log SRm	***Low***	Thyroid, Gallbladder, Skin melanoma, Connective tissue,	Leukemia, Hodgkin Lymphoma, Eye,Other skin cancers,	Other Endocrine glands, Pharynx
	Medium	Brain, Colon, Multiple Myeloma, Pancreas, Non-Hodgkin Lymphoma, Rectum & Anus, Kidney,	Bone, Salivary glands, Small Intestine,	
	High	Stomach,	Liver, Bladder	Nose & Sinuses, Mouth, Lip, Tongue, Lung, Oesophagus, Larynx,

SRm_{ik} (Low): ≤ 0.158; SRm_{ik} (Medium): 0.159 to 0.344; SRm_{ik} (High): > 0 .344
SRv_{ik} (Low): ≤ - 0.755; SRv_{ik} (Medium): - 0.756 to 0.198; SRv_{ik} (High): > 0.198

Table 8-11: Cancer sites stratified according to low, medium and high levels of incidence rates, sex ratio magnitude (SRm) and variances (SRv) in 281 registries in 2003-2007 on log scale.

		Low incidence			Medium incidence			High incidence		
		Log SRv			*Log SRv*			*Log SRv*		
		Low	*Medium*	*High*	*Low*	*Medium*	*High*	*Low*	*Medium*	*High*
Log SRm	*Low*		Leukemia, Eye,	Pharynx,	Thyroid, Gallbladder, Connective-tissues,	Hodgkin-Lymphoma,		Skin-melanoma,	Other skin cancers	endocrine glands
	Medium		Bone, Salivary glands, Small Intestine,		Brain, Multiple-Myeloma, Pancreas			Colon, Non-Hodgkin Lymphoma Rectum & Anus, Kidney,		
	High			Nose & Sinuses, Lip			Oesophagus, Larynx	Stomach,	Liver, Bladder	Mouth, Tongue, Lung

Low incidence: ≤ 1.63 / 10^5; Medium incidence: 1.64 to 6.52 / 10^5; High incidence: > 6.52 / 10^5
SRm (Low): ≤ 0.52; SRm (Medium): 0.53 to 0.61; SRm (High): > 0.61
SRv (Low): ≤ 0.003); SRv (Medium): 0.004 to 0.014; SRv (High): > 0.014

Cancers with low SRm and SRv:

In all three time periods (Tables 8-6, 8-8, and 8-10), cancers of the *thyroid* and *gallbladder* had *low SRm and low SRv,* signifying their stability as the two most female-dominant cancers worldwide. *Skin melanoma* was also a relatively stable female dominant cancer, although SRv was relatively higher in the earliest period. Cancers that were identified as having *low SRm and SRv* (in Tables 8-6, 8-8, and 8-10) remained as *low SRm and SRv* even when these cancers are stratified according to low, medium, and high incidence levels (Tables 8-7, 8-9 and 8-11). Worldwide, cancers of the *thyroid* and *gallbladder* had neither very low nor very high incident rates, whereas *skin melanoma* and *colon* had high incidence rates. Kidney cancer has low variance of sex ratio in all time periods, and the incidence has increased in the most recent time periods.

Cancers with high SRm and SRv:

High SRm and SRv are suggestive of highly variable patterns of exposures across different geographic regions. During 1973-77, 1988-93, and 2003-07, cancers with a high SRm and SRv include *lung*, *larynx*, and *esophagus*. Cancers of the *pharynx* and *liver* also had the same pattern on log-transformed version of magnitude and variance. *Bladder* cancer showed stability, and was male dominant with a highly variable SR in the two more recent time periods. *Lung, liver* and *bladder* cancers had high incidence rates worldwide. Cancer of the *esophagus* was stable in terms of its incidence rate across time. Cancers identified in this category were further described in terms of smoking prevalence in males and females for the existence of possible gender-bias in cancer registration (Section 8.5).

Other patterns of SRm and SRv:

Cancers that remain in the middle of 3×3 matrix of SRm and SRv with some constancy through time are *Hodgkin's lymphoma*, and low incident cancers such as cancer of *salivary glands* and *small intestine*. A consistent low levels of SRv is observed for non-*Hodgkin lymphoma*, cancers of the *rectum and anus* as well as *pancreas*.

8.3 Time trends and intra- & inter- regional variability of SR

This section presents findings using the methods described in Section 7.6. Here we present results from mixed effects regression analyses conducted to examine intra- and inter-regional variability of SR across time using cancer registries worldwide. The SR of 21 different types of cancers from long-standing registries were analyzed under three different scenarios (Appendix 6: A 6-a to A 6-u). From the Tables in Appendix 6, we chose the random-intercept & random-slope models (RI & RS Model *d*s) because overall, they had the lowest modelling error and were robust enough to deal with departures from the assumption of independence of errors or residuals in the models. The results from three RI & RS models for each cancer, in each of three scenarios are condensed in Tables 8-12 to 8-14. In this thesis, most of the discussion relates to the findings in Table 8-14, which are based on changes in SR using the same registries (N=76) for each year from 1983-2007.

SR is a *repeated-measures* variable (i.e., there are multiple measures of the same outcome variable (SR) on the same unit of observation (registry)). For example, in scenario C, SR is measured 25 times from 1983 to 2007 in each registry. Therefore, it is the registry that is measured multiple times. As SR is measured repeatedly on a registry population, these observations are not independent. For example, the SR in 1953 from one specific Italian registry is very likely to be related to the SR in 1954, 1955 and the following years in that

same registry. Because SR observations are not independent, the errors (or residuals) in general linear models will not be independent, which is a critical assumption in this type of model (263). If this assumption is violated, then the residual variance is under- or over-estimated and p-values are inaccurate. Repeated measures may also have non-constant variance over time and over registries.

Table 8-12: Results of regression models (Random-intercepts & Random-slopes) of 21 types of cancers for registry-regions in 1953-2007 (Scenario A: ranging from 3 to 113 registries in different years)*

Cancers	βo	β_1	*SE* β_1	(95% CI)	intra- regional variance (δ^2)	inter-regional variance (τ_1^2)	Variation in slopes (τ_{12})
Larynx	21.27	-0.193	0.054	(-0.299, -0.086)	130.5	838.44	0.22
Lung	11.89	-0.173	0.018	(-0.028, -0.137)	3.45	114.64	0.03
Esophagus	8.15	-0.052	0.024	(-0.099, -0.005)	30.45	146.78	0.04
Oral cavity & pharynx	6.88	-0.067	0.012	(-0.090, -0.042)	3.52	46.58	0.01
Bladder	5.04	-0.005	0.006	(-0.017, 0.008)	3.23	7.74	0.002
Liver	3.24	0.002	0.006	(-0.009, 0.013)	4.53	8.13	0.001
Hodgkin lymphoma	2.47	-0.021	0.003	(-0.027, -0.014)	1.61	1.05	0.0003
Stomach	2.16	0.002	0.002	(-0.001, 0.005)	0.36	0.1	0.001
Pancreas	2.07	-0.01	0.002	(-0.016, -0.009)	0.29	0.45	0.0002
Kidney	2.02	0.004	0.002	(-0.001, 0.008)	1.09	0.17	0.0001
Non-Hodgkin lymphoma	1.79	-0.006	0.001	(-0.008, -0.003)	0.37	0.73	0.000002
Leukemia	1.68	-0.002	0.001	(-0.009, -0.002)	0.28	0.05	0.000005
Bone	1.50	0.004	0.003	(-0.001, 0.010)	2.71	0.07	0.00001
Multiple myeloma	1.49	0.015	0.002	(-0.002, 0.005)	0.71	0.13	0.00002
Rectum & anus	1.47	0.005	0.001	(0.003, 0.007)	0.16	0.1	0.0001
Brain	1.34	0.004	0.006	(-0.007, 0.015)	0.85	4.44	0.003
Eye	1.26	0.007	0.004	(-0.0005, 0.0138)	2.32	0.29	0.0004
Colon	1.06	0.007	0.001	(0.004, 0.009)	0.08	0.16	0.0008
Skin melanoma	0.94	0.004	0.002	(-0.001, 0.009)	0.52	1.05	0.0003
Gallbladder	0.64	0.008	0.002	(0.005, 0.115)	0.38	0.17	0.0001
Thyroid	0.53	-0.004	0.0005	(-0.005, -0.003)	0.06	0.01	0.000003

Results are summarized from appendix 6

*SR is the outcome variable with calendar year as independent variable.

Table 8-13: Results of regression models (Random-intercepts & Random-slopes) of 21 types of cancers for registry-regions in 1983-2007 (Scenario B: 76 - 113 registries in each year).*

Cancers	βo	β_1	*SE* β_1	(95% CI)	intra- regional variance (δ^2)	inter-regional variance (τ_1^2)	Variation in slopes (τ_{12})
Larynx	22.08	-0.211	0.071	(-0.350, -0.071)	132.51	1301.22	0.39
Lung	11.71	-0.169	0.017	(-0.203, -0.134)	3.36	113.44	0.03
Esophagus	8.73	-0.066	0.022	(-0.109, -0.022)	23.26	129.48	0.03
Oral cavity & pharynx	7.44	-0.079	0.013	(-0.105, -0.053)	3.53	55.44	0.02
Bladder	5.28	-0.101	0.007	(-0.024, 0.004)	3.59	8.33	0.002
Liver	3.67	-0.01	0.01	(-0.021, 0.006)	4.8	6.14	0.0004
Hodgkin lymphoma	2.63	-0.024	0.005	(-0.034, -0.014)	1.69	2.73	0.001
Stomach	2.36	-0.002	0.002	(-0.01, 0.002)	0.39	0.09	0.0001
Pancreas	2.06	-0.01	0.002	(-0.015, -0.009)	0.27	0.4	0.0001
Bone	2.02	-0.007	0.01	(-0.027, 0.014)	3.13	0.001	0.009
Kidney	1.89	0.006	0.003	(-0.0001, 0.122)	---	---	---
Leukemia	1.85	-0.0106	0.002	(-0.0009, -0.003)	0.29	0.98	0.00001
Non-Hodgkin lymphoma	1.83	-0.006	0.005	(-0.015, 0.003)	0.39	---	---
Multiple myeloma	1.53	0.001	0.003	(-0.005, 0.007)	0.69	0.001	0.002
Rectum & anus	1.51	0.004	0.001	(0.001, 0.007)	0.17	0.01	0.00003
Brain	1.39	0.003	0.006	(-0.009, 0.014)	0.97	4.83	0.003
Eye	1.18	0.008	0.005	(-0.0025, 0.0189)	2.55	0.29	0.0007
Colon	1.15	0.005	0.001	(0.002, 0.007)	0.29	0.08	0.0001
Skin melanoma	0.96	0.003	0.003	(-0.003, 0.009)	0.57	1.52	0.001
Gallbladder	0.59	0.01	0.002	(0.001, 0.013)	0.43	0.09	0.0001
Thyroid	0.54	-0.004	0.0001	(-0.005, -0.002)	0.04	0.03	0.00001

Results are summarized from appendix 6

*SR is the outcome variable with calendar year as independent variable.

Table 8-14: Results of regression models (Random-intercepts & Random-slopes) of 21 types of cancers for registry-regions in 1983-2007 (Scenario C: 76 registries in each year).*

Cancers	βo	β_1	*SE* β_1	(95% CI)	intra- regional variance (δ^2)	inter-regional variance (τ_1^2)	Variation in slopes (τ_{12})
Larynx	20.2	-0.193	0.079	(-0.349, -0.036)	132.01	1096.01	0.35
Lung	9.99	-0.141	0.017	(-0.174, -0.109)	1.57	69.42	0.02
Esophagus	7.83	-0.052	0.026	(-0.104, -0.0001)	22.24	132.53	0.03
Oral cavity & pharynx	7.28	-0.08	0.017	(-0.114, -0.045)	3.51	65.22	0.02
Bladder	5.09	-0.011	0.007	(-0.025, 0.002)	3.36	6.22	0.001
Liver	3.83	-0.009	0.008	(-0.025, 0.006)	5.4	5.74	0.0003
Hodgkin lymphoma	2.74	-0.027	0.005	(-0.038, -0.017)	1.57	2.62	0.001
Stomach	2.37	-0.002	0.003	(-0.007, 0.003)	0.44	0.1	0.0001
Pancreas	2.08	-0.013	0.002	(-0.017, -0.009)	0.21	0.53	0.0002
Kidney	1.89	0.005	0.003	(-0.002, 0.112)	0.75	0.39	0.0002
Bone	1.85	-0.005	0.006	(-0.017, 0.007)	2.59	0.62	0.0005
Leukemia	1.85	-0.006	0.002	(-0.009, -0.002)	0.24	0.12	0.00002
Non-Hodgkin lymphoma	1.76	-0.006	-0.002	(-0.009, -0.002)	0.17	0.21	0.0001
Brain	1.65	-0.004	0.002	(-0.008, 0.001)	0.51	0.17	0.00004
Rectum & anus	1.53	0.004	0.002	(0.001, 0.007)	0.14	0.16	0.0001
Multiple myeloma	1.52	0.0005	0.002	(-0.004, 0.005)	0.45	0.23	0.0001
Colon	1.21	0.003	0.002	(-0.0002, 0.006)	0.08	0.31	0.0001
Eye	1.2	0.001	0.006	(-0.003, 0.021)	2.33	0.57	0.001
Skin melanoma	0.92	0.005	0.003	(-0.018, 0.012)	0.53	1.14	0.0001
Gallbladder	0.65	0.008	0.002	(0.004, 0.012)	0.38	0.16	0.00004
Thyroid	0.54	-0.004	0.001	(-0.005, -0.003)	0.04	0.03	0.00001

Results are summarized from appendix 6

*SR is the outcome variable with calendar year as independent variable.

In light of the limitations of general linear and fixed-effect models, the most viable analytical approach for the current study, therefore, was mixed-effects modelling on repeated measures data. This approach allows for the non-independence of data within registries, as well as potentially non-constant variance.

As mentioned in the preceding paragraph, the unit of analysis in this study is 'cancer registry' representing a certain population or group of populations. We did however, analyze all registries individually and compare them under different scenarios (Model *b* in Appendix 6: A 6-a to A 6-u) before moving to models *c* and *d*. The total error was reduced for all cancers when using model *b* compared to model *a*. However, based on the values of the AIC, model *d* was selected for interpretation of results of all cancer types (Tables 8-12 to 8-14).

Based on the results (Appendix 6), the best fitting models for most cancers was model *d* (i.e., RI & RS model). The exception was *thyroid* cancer, which showed a better fit with a random-intercept only model (i.e., the AIC was lower in the RI model *c* compared to the RI & RS model *d* (Appendix 6; A 6-u)). This was consistent within Scenarios A, B and C. A slightly higher AIC was noted for cancers of *gallbladder* (Appendix 6; A 6-t) and *bone* (Appendix 6; A 6-j). For *brain* cancer, both models *c* and *d* were a good fit because the AIC was the same in both (Appendix 6; A 6-p). Because of inherent advantages, the results of the RI & RS models for 1983-2007 (Scenario C) were preferred for interpretation of the study results.

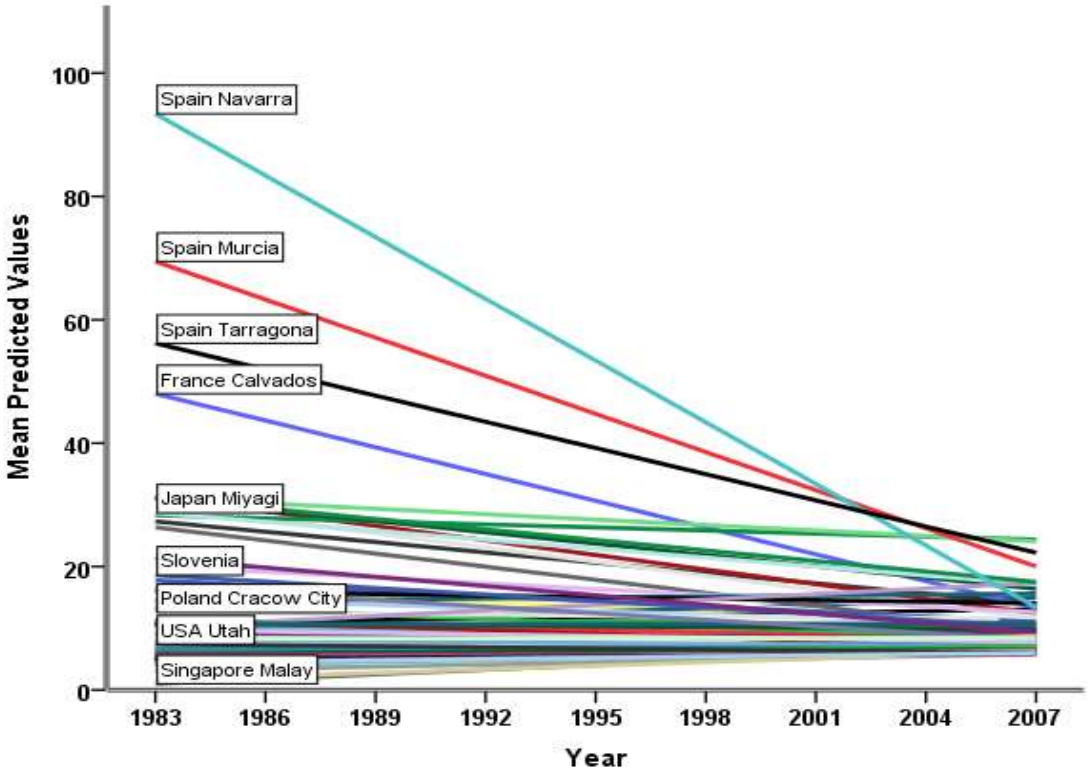
The RI & RS models estimated the average linear annual rate of change in SR (β_1), either a decrease or increase, over 25 years (1983-2007) by taking inter-regional variations into account (intercepts: τ_1^2; Appendix 6: A 6-a to A 6-u). Confidence levels are also presented

for each of the β_1 for all cancers. An added advantage of RI & RS models was that the variations in SR trajectories of an individual registry were also modelled (known as slopes: τ_2^2 (appendix 6)). The RI & RS model also focused on intra-regional variation (δ^2). Unlike fixed-effects models, in RI & RS models, intercepts are examined to account for the fact that SR changes over time in different registries. Also RI & RS models provide flexibility in that SR trajectories (slopes) are allowed to change as well.

Table 8-14 provides a summary of the results for 21 cancers from 76 registries in 1983-2007. The SR in the baseline year (1983 (β_0)) is provided, as well as the annual rate of change over 25 years (β_1). All cancers in Table 8-14 are tabulated from highest to lowest SR in 1983. As indicated by βo, the SR was generally high for most cancers with a male predominance except for *thyroid*, *gallbladder*, and *melanoma of skin* where age-standardized incidence rates were higher in females. In 1983, the highest SR (i.e., 20.2 on average) across all 76 registries was noted for cancer of the *larynx,* and the lowest (0.54) was noted for *thyroid* cancer. The average annual rate of decrease in cancer of *larynx* was -0.193, which was one of largest decreases worldwide in 25 years and this was a significant decrease (95% CI: -0.349, -0.036). Compared to other cancers, *laryngeal* carcinoma had very high intra- and inter-regional SR variations. Also, the inter-regional SR variation was very high compared to intra-regional SR variations for laryngeal cancer. The results of the mixed-effects analyses, which are plotted for 12 cancers, provide data on the random intercepts and random slopes of different registries (Appendix 7; A 7-a to A 7-h). These figures also provide an overview of intra- and inter- regional SR variations for individual registries, and selected cancers are presented based on their intra- (δ^2) and inter- (τ_1^2) regional variations. Cancers where δ^2 was lower than τ_1^2 include *larynx, esophagus*, *lung*,

liver, and *Hodgkin's lymphoma*. *Non-Hodgkin's lymphoma* and *thyroid* represented cancers where δ^2 was almost equal to τ_1^2. *Gallbladder*, *leukemia*, *brain* and *stomach* had a higher δ^2 compared to τ_1^2. Cancer of the *eye* had a uniquely higher δ^2 compared to an inter-regional variation τ_1^2. The slopes for each registry, also termed "*trajectories",* represent the annual rate of change in SR for that cancer. For cancer of *larynx*, while variation in the trajectories across registries was highest, there was also a significant covariance between trajectories implying that registries with the highest SR at baseline, on average, had steeper declines over 25 years [$\tau_{12} = -19.19$]. As an example, Figure 8-7 shows trajectories for *laryngeal* cancer from 1983 to 2007. The top three trajectories in Figure 8-7 represent three regions in Spain (i.e., Navarra, Murcia, and Tarragona). These three registry populations had the highest SR at baseline in 1983 and a steeper decline compared to others. The registry from Calvados, France is among the trajectories with high SR on average at baseline and a very steep decrease over the two decades. The lowest baseline SR was observed for a registry in Singapore and its trajectory increased slightly in 25 years. A registry from Utah, US showed one of the most stable SR trajectories throughout 25 years. Depictions of the varying trajectories of SR, also signifying intra- and inter-regional variation of SR for different cancers are shown in Appendix 7: a brief description of results is provided for each Figure.

Figure 8-7: Laryngeal cancer: sex ratio magnitude (mean predicted value) plotted against calendar year from 1983-2007 depicting inter-registry-regional (τ_1^2) and intra-registry-regional (δ^2) variations of sex ratios.



RI & RS model based on 76 cancer registries where δ^2 (132.01) < τ_1^2 (1096.01).

When the RI-only model is compared to the RI & RS model in any given scenario (A, B, and C), the variation in random slopes (i.e., termed *trajectories* in the RI & RS model) brought changes in other parameters of model (for all cancers: Appendix 6). By including variation in trajectories (τ_2^2) in the model, the variation of random intercepts (inter-regional variations in SR) increased many-fold (e.g., for cancer of *oral cavity* and *pharynx*, $\tau^2 = 5.16$ in RI-only model increased to $\tau_1^2 = 65.22$ in RI & RS model in Scenario C (Appendix 6; A 6-d). At the same time, the intra-regional variation δ^2 decreased from 4.51 to 3.50. There were a few exceptions (e.g., stomach cancers did not show any change with the introduction of varying trajectories in the model when transitioning from RI-only to RI & RS

model)(Appendix 6; A 6-h). This is also evident in the graph shown in Appendix 7 (A 7-j), where the only trajectory that showed a change in SR was in Neuchatel, Switzerland in 1983-2007. The rest of the trajectories showed similar patterns of intra- and inter-regional SRv. *Leukemia* and *brain* cancers are the most interesting since, unlike all other cancers, the random intercepts or inter-regional variation declined with the introduction of variations in slopes, indicating that these two cancers are the most stable cancers, and that the SR did not vary over time (Appendix 6; A 6-l and A 6-p). Additionally, cancer of the kidney has the same stable pattern (data on trajectories not shown), where the average sex ratio at the baseline is almost 2, and the average annual rate of change is consistent from 1983 to 2007.

The results from the RI and RS models in 1983-2007 (Table 8-14) show that cancers can be categorized according to whether intra- and inter-regional variations in SR are small or large. For intra-regional variations in SR denoted by δ^2, the range of values can be divided into less than one, and more than or equal to one. Cancers with $\delta^2 < 1$, the values range from 0.04 (*thyroid*) to 0.75 (*kidney*), whereas cancers with $\delta^2 > 1$, values range from 1.57 (*lung*) to 132.01 (*larynx*). For cancers with less intra-regional variation, SR was very stable over the 25-year period. Cancers that affect either males or females predominantly appear to change during this period. The annual rate of change in SR increases and decreases but only to a small extent, (i.e., there is very little change from one year to next). For cancers with large intra-regional variations, the average rate of change indicates that the cancer incidence in males and females within a region change in different ways possibly depending on causes that affect them unequally within a region. These cancers in increasing order of intra-registry variations δ^2 include: *thyroid* (0.04), *colon* (0.08), *rectum & anus* (0.14), *non-Hodgkin's lymphoma* (0.17), *pancreas* (0.21), *leukemia* (0.24), *gallbladder* (0.38), *stomach*

(0.44), *multiple myeloma* (0.44), *brain* (0.51), *skin melanoma* (0.53), *kidney* (0.75), *lung* (1.57), *Hodgkin's lymphoma* (1.57), *eye* (2.33), *bone* (2.59), *bladder* (3.36), *oral cavity & pharynx* (3.51), *liver* (5.4), *esophagus* (22.24), and *larynx* (132.01).

For inter-registries denoted by τ_1^2, for each cancer, the range of values can also be divided into less than one, and more than or equal to one. These values representing variation of intercepts range from 0.03 (*thyroid*) to 0.62 (*bone*) for cancers where $\tau_1^2 < 1$, and for cancers where $\tau_1^2 > 1$, the range is from 1.14 in lung to 1096.01 for *larynx*. With the introduction of slopes or trajectories in RI & RS model, the inter-regional variations of SR signify the difference of the individual trajectory of a region (registry) from the overall trajectory of all regions (all registries). Some of the inter-regional variations are small such as *thyroid, stomach*, and *leukemia*, while some are quite big e.g., cancer of *lung, esophagus* and *larynx*. The cancers behave differently from region to region most likely because the causes of cancers behave differently in different populations of regions.

8.4 Cancers by gender inequality and smoking patterns

To explore the issue of potential gender bias due to the possibility of differential disparities created by health seeking behaviors such as access to health care facilities and therapeutic treatment of cancers, we selected cancers with high SRm and SRv (e.g., in 2003-07 from Table 8-10) that are known to be caused by tobacco smoking. Table 8-15 (a and b), presents cancer registries (selected from CI5-vol X) in countries listed according to low and high gender inequality index with age-adjusted prevalence estimates of smoking any tobacco product among males and females. Sex ratios for cancers of lung, bladder, esophagus and larynx are tallied with the two statistics retrieved from UN (gender inequality index) and WHO (tobacco prevalence).

Table 8-15 (a) shows world rankings of countries where the gender inequality is lowest, and as such it is likely that females have equal access to health care services in these countries. Based on the index of gender inequality, it can also be assumed that gender bias can be less of an issue in these cancer registries, as is evident from relatively similar values of sex ratios for four selected cancer types in Sweden, Denmark and Norway. The smoking prevalence in most of the countries in Table 8-15 (a) is somewhat similar in both sexes with the exception of Sweden where relatively more females are smokers, and vice versa in Czech Republic.

Table 8-15: Sex ratios of lung, bladder, esophageal and laryngeal cancers with country level prevalence estimates of smoking tobacco.

(a) Cancer registries in countries with lowest gender inequality index (GII)

Registry	GII (World Rank)[a]	SRm[b]				Smoking prevalence[c]	
		Lung	Bladder	Esophagus	Larynx	Male	Female
Sweden	0.053 (1)	1.2	3.6	3.5	6.3	19.60	24.50
Denmark	0.056 (3)	1.3	3.4	2.9	4.7	33.60	30.40
Norway	0.085 (5)	1.6	3.6	3.7	6.8	31.80	24.40
Finland	0.088 (6)	3.2	4.9	3.2	12.0	30.10	24.10
Belgium	0.104 (7)	3.4	5.0	3.8	7.9	46.40	40.10
Austria	0.118 (10)	2.5	3.9	6.0	8.6	36.40	30.90
Spain	0.118 (11)	7.1	14.9	10.7	28.2	26.10	26.60
Iceland	0.125 (12)	1.1	4.1	2.9	9.5	27.70	21.80
Czech R	0.153 (18)	3.7	3.7	6.6	12.8	31.10	17.90
N. Zeal	0.188 (29)	1.4	2.8	3.7	5.8	26.50	26.00

a. GII data from UN Human Development Report (2005 figures). Ranking is out of 160 countries
b. Data from CI-5 Vol X (2003-07).
c. Data from WHO Report on global tobacco epidemic of current smokers published in 2005 (Age-standardized prevalence estimates).

(b) Cancer registries in countries with moderate to high gender inequality index (GII)

Registry	GII (World Rank)[a]	SRm[b]				Smoking prevalence[c]	
		Lung	Bladder	Esophagus	Larynx	Male	Female
China	0.228 (35)	1.7	2.9	3.3	6.7	47.50	27.80
Bulgaria	0.251 (36)	6.6	4.7	6.3	25.3	26.10	2.90
Bahrain	0.331 (45)	2.8	4.6	2.2	17.5	51.00	1.90
Kuwait	0.374 (62)	2.4	2.3	1.3	3.0	38.30	1.40
Ecuador	0.454 (73)	1.3	2.5	3.1	6.0	36.10	30.60
Egypt	0.581 (109)	3.2	3.8	1.9	14.0	28.70	1.30
Uganda	0.592 (115)	1.0	1.4	2.7	12.5	20.90	3.20
India	0.619 (119)	2.7	3.5	1.5	6.8	33.10	3.80
S Arabia	0.672 (135)	2.2	4.3	1.2	17.0	25.60	3.60
Belarus	0.151 (145)	12.1	7.6	16.5	47.5	36.60	25.40
Qatar	0.579 (154)	4.7	4.0	0.6	21.7	20.20	3.10

d. GII data from UN Human Development Report (2005 figures). Ranking is out of 160 countries
e. Data from CI-5 Vol X (2003-07).
f. Data from WHO Report on global tobacco epidemic of current smokers published in 2005 (Age-standardized prevalence estimates).

Another prominent exception that is observed is an exceptionally high sex ratio of cancers when the prevalence of smoking is similar in males and females (relative on high side in females). Also, Spain's gender inequality index shows that it is a fairly gender balanced country in terms of perceived economic advantages and is ranked eleventh. Yet an extremely high sex ratios of lung cancer in particular is indicative that the male and female completeness of ascertainment (and other artifacts) might not have been similar in some of the Spanish cancer registries.

Table 8-15 (b) shows world rankings of countries where the gender inequality is modest (China, Bulgaria, Bahrain, Kuwait and Ecuador) and highest (Egypt, India, Saudi Arabia, Uganda and Qatar). There is a possibility that females might not have been recorded in the registries because they might not have the same level of access to health care services as the males in these countries. However, the sex ratios of four selected cancers (in particular for lung) in Egypt, India, and Saudi Arabia show similarities with Finland and Belgium. This is

in spite that there seems to be is a considerable gap among males and females in smoking prevalence in Egypt, India, and Saudi Arabia. Cancer registry from Qatar has the highest sex ratio for lung, bladder, and laryngeal cancer with the exception of esophageal cancer that shows the lowest ratio (0.6). Hence gender-bias can be possibility in a registry from Qatar.

Lung cancer still remains a male dominant cancer, despite the low gender inequality and similar prevalence of smoking among males and females in some of the well-established oldest cancer registries in Western European countries (Sweden, Denmark and Norway). For countries that rank highest in gender inequality, SR of lung cancer is relatively similar to countries with low gender inequality, with the exception of Qatar, where a very high SR for lung cancer is observed (4.7). Sex ratio for esophageal cancer is higher in registries belonging to developed countries with low gender inequality e.g., SR is 6.0 and 6.6 in Austria and Czech Republic. Cancer of larynx is higher in both registries belonging to countries that have low and high gender inequality.

Table 8-16 displays the longitudinal associations between gender inequality and sex ratios of cancers. Results of the mixed effects regression models are shown that was carried out on cancers of lung, bladder, esophagus and larynx. These are the cancers where smoking is thought to be major etiological factor. The gender inequality as assessed through UN's gender inequality index was categorized as low and high. In the model, both linear and quadratic trends were incorporated as well as the interaction between gender inequality and the calendar time (in years). Negative values of coefficients indicate that the sex ratio declined with time.

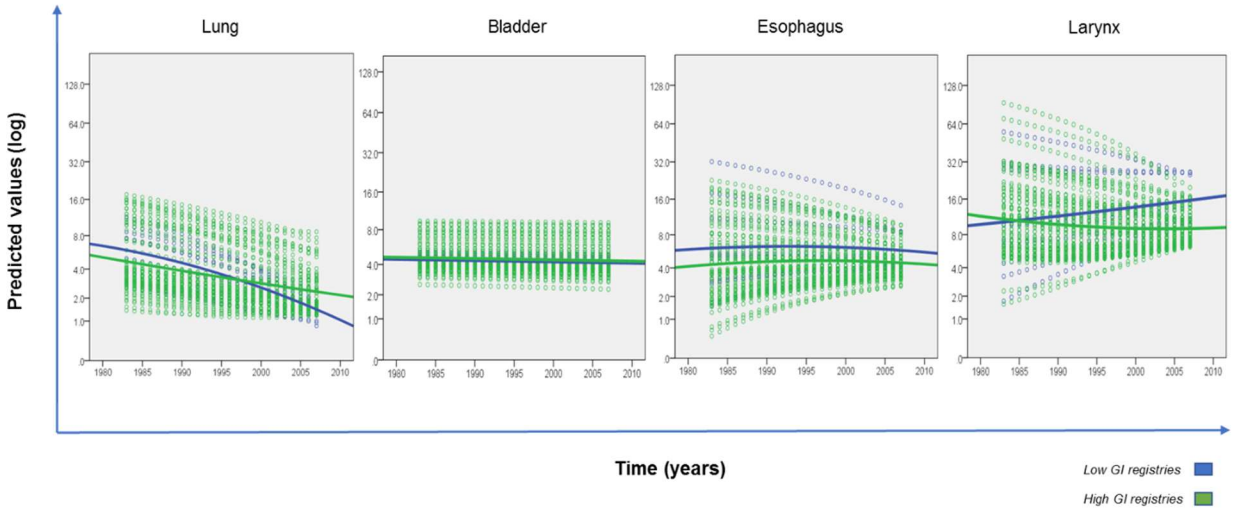
Figure 8-8 displays the results from table 8-16 graphically, in particular the interaction between gender inequality and time. These figures depict the predicted average sex ratio

changes between 1983 and 2007 in cancer registries from 32 countries defined according to two levels of gender inequality. In lung cancer, the sex ratio trajectories (slopes) shows downward trends for both levels of gender inequality (low and high) indicating that the incidence of lung cancer is increasing in females. Both trajectories converge in 1993 and thereafter decline with increased divergence as a function of sex ratio (i.e., at the rate of 0.085). The pattern of sex ratio trajectories of low and high gender inequality countries remain constant in bladder cancer, although the trajectory for low gender inequality countries are higher than the countries where the gender inequality is higher. For cancer of esophagus, both trajectories are converging whereas for larynx, opposite is true.

Table 8-16: Predicted average sex ratio changes between 1983 and 2007 in cancer registries from 32 countries defined according to two levels of gender inequality

	Lung			Bladder			Esophagus			Larynx		
	Coef.	**p**	**95 % CI**	**Coef.**	**p**	**95 % CI**	**Coef.**	**p**	**95 % CI**	**Coef.**	**p**	**95 % CI**
Fixed effects												
Intercept	15.62	<0.001	9.67,21.56	3.00	0.119	-0.77,6.77	10.75	0.070	-0.88,22.38	26.03	0.101	-5.14, 57.20
Time	-0.346	<0.001	-0.48, -2.07	0.082	0.306	-0.07,0.24	-0.231	0.915	-0.445,0.399	-4.888	0.376	-1.569,0.593
$Time^2$	0.002	0.014	0.0003,0.002	-0.001	0.201	-0.003,0.001	-0.001	0.656	-0.006,0.004	0.005	0.327	-0.006, 0.018
GI	-3.391	0.259	-9.34,2.55	0.086	0.942	-2.26,2.43	-5.402	0.235	-14.398,3.592	4.802	0.709	-20.71, 30.320
GI*Time	0.085	0.092	-0.01,0.18	0.007	0.761	-0.04,0.05	0.069	0.385	-0.089,0.227	-0.224	0.349	-0.70, 0.2
Random effects												
Residual variance	1.561	<0.001	1.46,1.67	3.355	<0.001	3.14,3.59	22.248	<0.001	20.774,23.806	131.992	<0.001	123.11, 141.5
Intercept variance	68.204	<0.001	48.90,95.11	6.230	0.001	3.52,11.02	129.442	<0.001	86.668,193.33	1089.645	<0.001	741.29,1601.66
Covariance	-1.099	<0.001	-1.47, -0.72	-0.068	0.030	-0.13, -0.07	-2.021	<0.001	-2.921, -1.121	-18.993	<0.001	-26.76, -11.22
Trend variance	0.182	<0.001	0.01,0.03	0.001	0.089	0.0003,0.0033	0.032	<0.001	0.019,0.053	0.338	<0.001	0.21, 0.52

Figure 8-8: Sex ratios average trajectories based on random intercept and random slopes models for registries with low and high gender inequality between 1983 and 2007; GI:gender inequality



Hence in this analysis, the examination of gender inequality and its lack of association with sex ratios suggest on one hand, that there seems to be no significant bias among the cancer registries in terms of differences in ascertainment of cancer cases in males and females, in few selected types of cancers. On the other hand, the direction of sex ratio trajectories of low and high gender inequality countries provides some clues on the etiology of these cancers in terms of smoking in males and females.

8.4 Spatial autocorrelations: Global Moran Indices

The statistical approaches in our study assumed that the outcomes (i.e., SR in each year) were not independent. In spatial data, it is often the case that some or all outcomes exhibit spatial autocorrelation. This occurs when the relative outcomes of two or more contiguous regions or populations are related to their distance. In the current study, the contiguous populations of registries in our data could be correlated and hence is likely to have similar

environmental, non-environmental, and socioeconomic conditions. Hence, when analyzing spatial data, it is important to check for autocorrelation.

In this section, findings are presented related to the spatial patterns of SR by taking the geographic distances between the populations covered by the registries into account. The contiguous geographic regions from cancer registries covered 102 locations in Europe and 52 locations in Asia that reported incidence rates in the 2003-2007 (CI5-Vol X). Global Moran's Index (I) measured the similarity in age-standardized incidence rates among males and females as well as SR in areas that are close geographically. This technique of spatial autocorrelation aimed to describe the scale over which the spatial patterns of SR of incidence occur, and that in turn can provide suggestions concerning the putative causal agents bringing about the patterns. If there is no spatial dependence, then Moran's I will be close to 0 indicating a random pattern, while values close to one indicate a clustered geographic pattern. Negative spatial autocorrelation from Moran's I indicate that neighboring values are dissimilar suggesting inverse spatial dependence or a dispersed pattern (see also Section 7.7.1). Therefore this approach presents results on the distribution of cancer incidence in terms of their SR for 28 cancers in Europe and Asia. These two continents were selected because of the availability of incidence rates from most regions and their geographical contiguity.

Tables 8-15 and 8-16 presents values of Moran's I across all 28 cancers in Europe and Asia respectively. All cancers are presented with rankings according to these values, indicating the pattern of clustering in geographically contiguous areas. The focus of spatial analysis is SR. However, Moran's I was also computed for incidence rates in males and females separately to have a more complete understanding of the spatial trends.

Table 8-17: Moran's I values with ranking of spatial clustering for sex ratios of 28 cancers in Europe

Cancers*	Moran's I					
	Male Incidence		Female Incidence		Sex Ratio	
	Value	*Rank*	*Value*	*Rank*	*Value*	*Rank*
Lip	0,0806	11	0,0644	11	**0.0124**	**12**
Tongue	0,1230	8	0,0286	15	**0.0824**	**6**
Mouth	0,1315	6	-0,0214	24	**0.0851**	**5**
Pharynx	0,0354	17	0,0276	16	**0.0192**	**11**
Oesophagus	0,0672	15	0,2852	1	**0.1156**	**1**
Stomach	0,0791	12	0,0958	7	**0.0955**	**4**
Small Intestine	0,0790	13	-0,0027	20	**-0.0082**	**20**
Colon	0,0509	16	-0,0417	25	**-0.0213**	**23**
Rectum Anus	-0,0125	25	-0,1499	28	**0.0364**	**8**
Liver	0,1868	2	0,1916	2	**0.1155**	**2**
Gall bladder	0,1313	7	0,0949	8	**0.0073**	**14**
Pancreas	0,1177	9	0,1458	4	**0.0745**	**7**
Nose Sinuses	0,0337	18	0,0203	18	**0.0329**	**9**
Larynx	-0,0049	24	-0,0008	19	**0.0056**	**15**
Lung	0,0268	19	0,0240	17	**-0.0049**	**18**
Bone	0,0109	21	0,0631	12	**-0.0032**	**17**
Multiple myeloma	0,0259	20	0,0819	9	**-0.0564**	**27**
Melanoma Skin	0,1684	4	0,0431	13	**-0.0959**	**28**
Connective tissue	0,0084	23	-0,0097	22	**-0.0064**	**19**
Salivary gland	-0,0247	26	-0,0130	23	**-0.0547**	**26**
Eye	-0,0747	28	-0,0841	27	**-0.0255**	**24**
Thyroid	0,1997	1	0,1115	6	**0.0223**	**10**
Bladder	0,0909	10	0,0673	10	**-0.018**	**22**
Hodgkin Lymphoma	0,1344	5	0,1511	3	**0.0088**	**13**
Kidney	0,0675	14	0,0389	14	**0.1142**	**3**
Brain	0,0097	22	-0,0069	21	**-0.0159**	**21**
Non Hodgkin Lymph	0,1719	3	0,1243	5	**0.000069**	**16**
Leukemia	-0,0649	27	-0,0449	26	**-0.0274**	**25**

*Cancers sorted according to ICD-10-O

Table 8-18: Moran's I values with ranking of spatial clustering for sex ratios of 28 cancers in Asia.

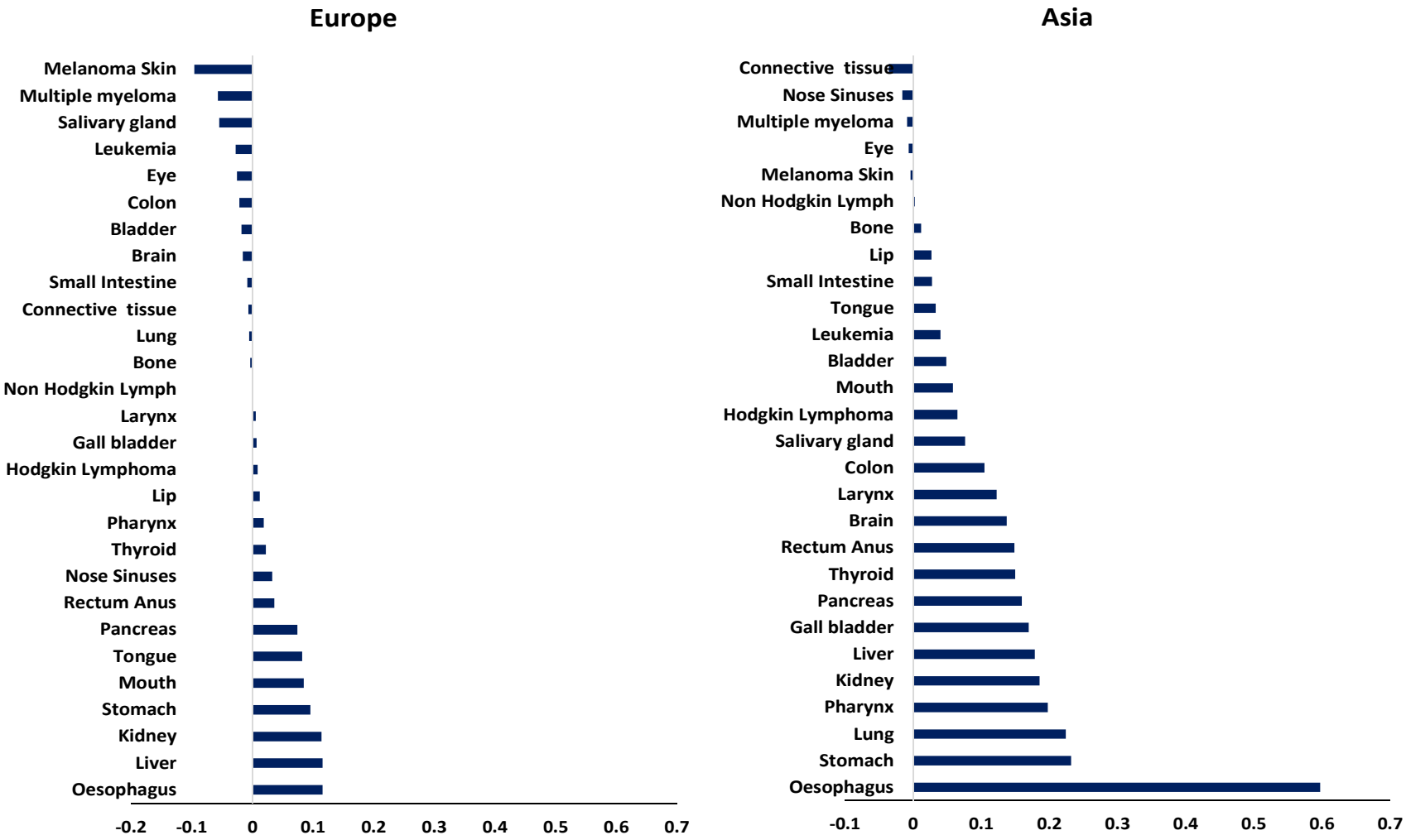
Cancers*	Moran's I					
	Male Incidence		Female Incidence		Sex Ratio	
	Value	*Rank*	*Value*	*Rank*	*Value*	*Rank*
Lip	0.1727	19	0.1389	17	**0.0272**	**21**
Tongue	0.2889	12	0.3505	8	**0.0332**	**19**
Mouth	0.4296	6	0.3872	6	**0.0586**	**16**
Pharynx	0.1097	21	0.1428	16	**0.1977**	**4**
Oesophagus	0.0181	26	0.0185	27	**0.5974**	**1**
Stomach	0.2554	13	0.2329	11	**0.2321**	**2**
Small Intestine	0.4137	8	0.3682	7	**0.0279**	**20**
Colon	0.4187	7	0.3185	9	**0.1048**	**13**
Rectum Anus	0.5888	2	0.5414	2	**0.14913**	**10**
Liver	0.3098	11	0.2002	15	**0.1785**	**6**
Gall bladder	0.6613	1	0.5113	4	**0.1699**	**7**
Pancreas	0.5752	3	0.5222	3	**0.1599**	**8**
Nose Sinuses	0.0297	24	0.0615	24	**-0.0156**	**27**
Larynx	0.1840	18	0.0891	23	**0.1230**	**12**
Lung	0.4080	10	0.2511	10	**0.2241**	**3**
Bone	0.1209	20	0.1092	19	**0.0123**	**22**
Multiple myeloma	0.1062	22	0.0933	22	**-0.0087**	**26**
Melanoma Skin	0.2273	15	0.2168	13	**-0.0036**	**24**
Connective tissue	0.0233	25	0.0295	26	**-0.0357**	**28**
Salivary gland	0.0072	28	0.0336	25	**0.0764**	**14**
Eye	0.0117	27	0.1051	20	**-0.0062**	**25**
Thyroid	0.4576	5	0.6282	1	**0.1499**	**9**
Bladder	0.4089	9	0.2049	14	**0.0490**	**17**
Hodgkin Lymphoma	0.2185	16	0.1018	21	**0.0653**	**15**
Kidney	0.4734	4	0.3975	5	**0.1858**	**5**
Brain	0.2098	17	0.0081	28	**0.1377**	**11**
Non Hodgkin Lymph	0.2478	14	0.2201	12	**0.0025**	**23**
Leukemia	0.0322	23	0.1388	18	**0.0402**	**18**

*Cancers sorted according to ICD-10-O

In Europe, Moran's I statistic for incidence rates ranges from 0.199 for *thyroid* cancer to – 0.074 for Non-Hodgkin's lymphoma in males and from 0.285 for *esophagus* to -0.149 for cancers of *rectum and anus* in females (Table 8-15). Therefore the highest clustering for incidence rates in Europe is found for *thyroid* cancer in men whereas for females, the clustering is highest for *esophageal* cancers. The greatest spatial clustering that is common to both males and females is *liver* cancer and to some extent *Hodgkin* and *non-Hodgkin's lymphoma*. However the values of Moran's I in all cancers in Europe is lower than 0.3. For cancers with high spatial clustering, one would expect to see regions with hot spots on a map of Europe. High clustering also means that the cancers tend to cluster in certain regions with high incidence and also band together in regions with low incidence. Therefore there will be clear demarcation of high incidence and low incidence spots on a map for highly clustered patterns.

Moran's values were also computed for SR of cancer incidence in Europe (Table 8-15). These are presented in Figure 8-8 where Moran's I can be visualized on a scale of -1 and +1. The ranking ranged from a very similar clustering of 0.115 (*esophagus*), 0.115 (*liver*), and 0.114 (*kidney*) to slightly negatively correlations such as -0.09 for *skin melanoma*. Moran's value for *esophageal cancer* is ranked one for SR in both males and females. Cancer of the *liver* is the only one that shows the same ranking of Moran's value for incidences in males and females, as well as for SR (ranked no. 2).

Figure 8-9: Spatial correlations- Moran's Index for Sex Ratios of cancer incidences in Europe and Asia.



Compared to Europe, regions in Asia depict a different spatial pattern as seen in Table 8-16 (i.e., the scale on which Moran's values are based shows very high clustering for incidence rates (e.g., the maximum value is 0.661 for males (*gallbladder*) and 0.628 (*thyroid*) in females)). In Asia, Moran's I statistic ranges from high values for gallbladder to 0.0072 (*salivary glands*) and 0.011 (*eye*) in males. For females, the values range from highest in *thyroid* cancer to 0.008 for cancers of *brain*. A tendency to high spatial clustering, (i.e., values at or higher than 0.5 in both males and females) are seen for cancers of *gallbladder*, *thyroid*, *pancreas*, *kidney* and cancers of *rectum and anus*. In females, spatial clustering for *Hodgkin* and *non-Hodgkin's lymphoma* in Asia shows similar pattern as in Europe. However, clustering seems to be high for these cancers in males in Asia. *Liver* cancer has the same values of Moran's I in Asia as in Europe for both males and females.

For cancers with high spatial clustering, one would expect to see regions with hot spots on a map. Previously, Rosenberg *et al*., showed that cancers tend to cluster in regions with high incidence and also to group together in regions with low incidence (254). Therefore, one would expect to have a clear separation of high and low incidence spots on a map for clustered patterns. *Leukemia* in both males and females shows a lack of spatial correlation.

One of the most striking findings in Asia is that, while cancer of the *esophagus* has one of the lowest values of Moran's I in males and females (0.018 in both, and ranked 26 and 27 out of all cancers), the SR for this cancer shows spatial clustering that is exceptionally high (about 0.6)and ranked one. The rest of cancers in Asia are less than 0.3 on Moran's scale. Non-Hodgkin's lymphoma and skin melanoma seems to be spatially randomly distributed.

Chapter 9 Discussion

Cancer surveillance has a long tradition of influencing cancer control policy, at both national and international levels (278). Comparison of incidence trends across regions worldwide is highly informative, and provides subtle clues about the causes of cancer and variation in the causes. The current study provides a fertile ground for speculating on causes and facilitates positing tentative hypotheses with regard to changing or stable patterns of cancer incidence and their sex ratios.

In the literature, most types of cancers show differences in incidence rates across populations, and depending on many causative factors, known or unknown, these rates change over time (279). The etiological role of environmental, genetic, and lifestyle factors in cancer causation is of course, of interest to cancer researchers. Doll and Peto, in their landmark study on the causes of cancers, concluded that differences in cancer rates are attributable to behavioral factors such as smoking, diet, reproductive and sexual behaviour (72). Their conclusion was viewed as a way forward at a time when epidemiological research on occupational exposures was making inroads in understanding etiology. Epstein and Swartz, however, criticized Doll and Peto's approach as being too focused on behavioral rather than looking onto other causes of cancers as well (280). These perspectives were however, largely unifying in terms of broadening our understanding of the complexities of cancer etiology.

Hence, incidence rates in males and females typically fluctuate or they can remain stable across time and geography, as evidenced in this current study (Figures 8-1 to 8-3). For this thesis, the complementary views of Doll & Peto (72) and Epstein & Swartz (280), and more

recently, Tomasetti & Vogelstein (281) on variation of cancer rates, were motivation in itself to explore incidence trends of different cancer types. These perspectives provided an important starting point for our study on the sex ratios of incidence. Most of the studies on incidence rates of cancers have reported that the rates are higher in males than females for a particular type of cancer, yet there are very few studies where consistency (or inconsistencies) of patterns or trends of sex ratios are used to raise a question or suggest hypotheses. It is only recently that there is a renewed interest in demonstrating the utility of sex ratios as a method that can be useful in inferring on possible or potential causes of cancers. Hence, as demonstrated by Trojano *et al* (33)., and Orton *et al* (36)., the utility of sex-ratio-methodology to suggest hypotheses on causes of multiple sclerosis, the present study also presents the same methodology to compare trends in cancer incidence across time and geography. The main premise of the analyses (as mentioned in Chapter 6), is based on assumption that if ascertainment of cancer cases among males and females in different registries, is similar, then SR is a better statistic to compare different populations represented by these registries. This is also because we believe that reported incidence rates are flawed indictors of true incidence rates and hence descriptive statistics based on comparisons of incidence rates are likely to be biased.

Discussion of the results follows in next sections, leading to dialogue based on the inductive template of inferences i.e., from basic observations, to delineating trends, presenting tentative hypotheses and then theorizing point of view and generate questions.

9.1 General observations on sex ratios of cancer incidence and existing literature

Making observation(s) is an integral starting point for any scientific inquiry. How the observations are collected, classified, interpreted, and used as the basis of theorizing is what a scientific inquiry is about (282). Hence in this section we review some of the studies on the SR of cancer incidence as well as mortality and survival based on what other investigators observed, and whether these complement observations from the current study.

In their 1959 publication, Kirchoff and Rigdon reviewed the literature on the SR of *lung* cancer (283) and reported that as early as 1871, Walshe (284), observed that *lung* cancer was more frequent in men. A summary analysis of cancers in ten metropolitan areas in the US in 1947-1948 (285) showed SR for the following cancers: *larynx* (12.33), *lip* (7.57), *pharynx* (5.38), *esophagus* (4.47), *lung and bronchus* (4.49), *tongue* (3.22), and *bladder* (1.19). Despite their careful statistical work, the results were no more accurate than results obtained through voluntary reporting by physicians and hospitals, supplemented by death certificates (286). The data from Atlanta, Birmingham, Dallas, New Orleans, San Francisco, Denver, Chicago, Detroit, Philadelphia, and Pittsburgh were combined into three geographic regions (North, South, and West) which masked statistical differences between cities.

The high SR (i.e., male dominance) in the incidence of *lung* cancer constituted an obvious line of inquiry into the etiology of lung cancer (287). Sex differences observed in early studies were investigated in later studies with an emphasis on smoking habits and consumption of alcohol (109, 288, 289). At a time when the role of smoking in the etiology of *lung* cancer was newly appreciated in 1950s and 1960s, and even later, very few studies reported a SR analysis that included all types of cancers, in addition to cancer.

Flamant *et al*. reported, for the first time, SR in a large number of cancer cases in a hospital setting in France (289). The SR was defined as the number of male cancer cases relative to females, and the age adjustment was undertaken using the population of France in 1956 (no further details were provided on the age adjustment). Their findings on 64,834 cases suggested a correlation between use of tobacco/alcohol with SR of more than 2.5. Sex ratio in various primary sites and subsites were: l*arynx* (37.5); *oral cavity* (24.5); *esophagus* (20.4); *lung* (13.2); *stomach* (5.0); *bladder* (3.5), *nasal cavities and sinuses* (2.2); *kidney* (1.8); *intestine and rectum* (1.3); and *skin* (1.2). Cancers of the *nervous system and endocrine glands* had an SR of 1.1. Their results are notable when compared to current study's data on France from 1975-2007 (Table 9-1) (e.g., the order of most cancers is similar with the exception of an increased SR in *bladder* cancer (7.2) and a 33% decrease for *skin cancer* (0.8)). However, this study can be critiqued (as the authors acknowledge) on the basis that it was conducted among patients in hospitals, and therefore the study population cannot be considered to be a representative sample. There was no clarity on the extent of geographical regions from where these patients belonged since the authors did not report the number of location of these French hospitals. Moreover, men and women attended the clinics in unequal numbers and the accuracy of site of the primary tumor location and ascertainment of diagnosis were questionable. Despite the limitations and the inadequate description of the SR, this study provided clues on the role of tobacco and alcohol in the risk of several cancers.

Table 9-1: Comparisons of sex ratio analysis of incidence and mortality with different studies for some of the cancer types

	Flamant, 1964[a] France 1943-1960 Incidence	*La Vecchia, 1988*[b] Switzerland 1951-1984 Mortality	*Cook, 2009*[c] US 1975-2004 Incidence	*Cook, 2011*[c] US 1977-2006 Mortality	*Raza, 2017*[d] France† 1975-2007† Incidence	*Raza, 2017*[d] Switzerland‡ *1975-2007* Incidence	*Raza, 2017*[d] US 1975-2007 Incidence	*Raza, 2017*[d] World 1953-2007 Incidence
Larynx	37.5	-	5.2	5.4	22.8	11.0	6.1	12.8
Oral cavity	24.5	-	3.1	3.1	13.1	4.4	3.0	4.0
Esophagus	20.4	11.0	3.5	4.1	9.0	8.1	4.4	5.6
Lung	13.2	11.1	2.1	2.3	8.9	4.2	2.4	4.6
Stomach	5.0	2.3	2.2	2.1	2.7	2.5	2.2	2.2
Bladder	3.5	3.6	3.9	3.4	7.2	4.8	3.9	4.7
Kidney	1.8	1.8	2.1	2.2	2.4	2.6	2.2	2.1
Skin	1.2	1.6	1.4	2.3	0.8	1.0	1.3	2.1
Brain	1.1	1.4	1.5	1.5	1.7	1.8	1.6	1.5

Databases used:

[a] Institut national d'hygiène, Paris, France

[b] Swiss Federal Office of Statistics

[c] Surveillance Epidemiology and End Results Program, NIH, USA

[d] Cancer Incidence in Five Continents, International Agency for Research on Cancer, Lyon, France

† Regions of France included in analysis: Bas Rhin, Calvados, Doubs, Haut Rhin, Herault, Isere, Somme, Tarn.

‡ Regions of Switzerland included in analysis: Geneva, Graubunden & Glarus, Neuchatel, St. Gall Appenzell, Valais, Vaud.

Sex ratio trends in mortality were examined in Switzerland from 1951 to 1980 using data from the Vaud Cancer Registry (290). Age-standardized mortality rates were computed using a European standard population. The analyses showed that mortality for most types of cancers was persistently higher in males. The SR ranged from 1.2 for *intestines*, *skin, brain* to 2 for *stomach* or *pancreas*, up to 11 for *lung* and *esophagus* (Table 9-3). The authors showed that SR for lung cancer was increasing from the 1950's to the 1960's, but the ratio noticeably declined afterward, probably reflecting trends in smoking prevalence in subsequent generations of Swiss men and women.

Levi *et. al*., studied SR of mortality and survival in Europe (291). Their study indicated that SR in survival rates were smaller than in mortality, indicating that better survival in females was not the primary reason for the death rate observed in females. Additional reports suggested that cancers are disproportionate by sex in incidence, mortality, and survival (292-295). Cook *et. al*., carried out the first study to systematically examine SR in cancer mortality and survival using a large data from Surveillance Epidemiology and End Results Program (SEER) in US (38). The authors conducted the only large scale SR analysis of incidence rates in the US using the same data program (37). As shown in Table 9-1, their study on SR of mortality in 2011 mirrored SR in incidences in 2009. These SR figures suggest that sex differences in cancer mortality are the result of differences in cancer incidence and not survival, and hence, the suggested further etiological investigations. However, the mirror effects of SR of incidence and mortality implied that the case-fatality rates of cancers are similar in US (which might not be true for other regions). The results from our data for SR in the US are consistent with the results reported by Cook *et al*. in the same time period (Table 9-1).

There are few large-scale studies on SR of cancer incidence that include population-based registries worldwide on such a large scale. We analyzed temporal and geographical variability in SR in selected cross-sections of time over 55 years, as well as longitudinally spanning 65 years. A relatively recent analysis by Edgren *et al*. using CI5 data from 1998-2002 (39, 296) speculated that much of the observed disparity in SR results from endogenous sex-specific biological factors. They proposed that these factors are either determinants of disease or they modify susceptibility to exogenous oncogenic factors. Considering many of these observations, along with rigorous inquiry into the underpinnings of variability of SR in cancer, the current study provides the basis to suggest tentative hypotheses as well as theorize or generalize the findings as part of our inductive inferential reasoning. This will further advance our understanding of the gaps in causes of cancers in terms of male-female differential.

9.2 Point of view - overview of associations: established or not?

Investigations of cancer incidence trends have typically focused on one specific type of cancer in either a specific region or several regions (297-299), and generate hypotheses based on observations of changing SR patterns (300). As pointed out before, table 9-1 summarizes early SR studies of mortality and survival rates (290, 291) along with a more recent study by Cook *et al* (38). Comparative assessment of cancer causes have been carried out previously (301). Several potential causes and explanations are revisited or conjectured in relation to the associations of some of the putative and known causes with SR of cancer incidence in this section.

9.2.1 Temporal trends and geographic variations

Lung cancer, as expected, has the highest incidence across time in this study. A sex ratio decline was seen in the incidence which was relatively more noticeable in 2003-07 (Table 8-2). Highest and lowest SR also varied across diverse geographical regions ranging from European registries in earlier period, to Turkey and Tunisia in the latter period (Appendix 4). Except for several regions in India (Sikkim, and Mizoram), the SR was never less than one, indicating that there are regions where females have the same or more exposure than males. The northeastern states of India are known as the cancer capitals, where more females are tobacco chewers and biddi smokers than males (302, 303). The use of tobacco water is also traditional in some communities in Mizoram (304). The very high variations in SR of *larynx, esophagus, mouth* and *tongue* in our data likely reflect historical exposure to tobacco smoking, as observed in US-SEER data (37). The dramatic decrease of SR in the US was likely because of convergence of smoking habits between men and women (37, 305).

Also, in this study, the inter- and intra-regional variation for lung cancer was one of the highest (Tables 8-12 to 8-14), suggesting that while the variation of SR within individual regions is small, the pattern is variable across regions. This may relate to similar behavioral practices in regard to some exposure inside the populations of those individual registries, rather than across registries from other geographic regions. Both the SRm and SRv were high for *esophageal* cancer, which has a pattern similar to *lung, larynx* and other cancers of aero-digestive tract and *bladder*. The SR imbalance in *esophageal* cancer has previously been explained in relation to obesity (306), concentrated production of gastric acid in males (307), defective lower esophageal sphincter (308) and variable intra-abdominal pressures (309). In our study, the highest SR were observed in France, Hungary, Spain and Korea. The

lowest were observed in the northwest territories of Canada, Kuwait, Qatar, Colombia and India (Sikkim) across time (Appendix 4). In places where the SR is low and alcohol is not implicated in esophageal cancer, hypotheses can likely relate to geography, climate, soil and dietary habits (310).

The present study showed a highest SR in *laryngeal* carcinoma (Bas Rhin-France: 122 in 1973-77; Zaragoza-Spain: 177 in 1988-1992; and Cuenca-Spain: 100 in 2003-07). In recent time (2003-07), very high SR were also observed in China and Korea (Appendix 4). This cancer is known as a disease of men, and few studies have attempted to document differences across sex (311, 312). Alcohol and smoking can be implicated in the etiology (313), and an interesting hypothesis on the higher frequency of vibration of vocal cords has been proposed as protective in females. Faster mucocilliary clearance in women was discussed in a study by Yang et. al (314). Cancer of the *larynx* has been implicated with inhaled carcinogens in the workplace, but the evidence remains sparse (e.g., cement dust, petroleum derivatives, aromatic hydrocarbons, wood and metal dust etc (312, 315-318)).

The incidence of cancers of *rectum and anus* was high across time. The SR of 1.7 is one of the most stable and its variance is also one of the lowest (Table 8-2). Studies have suggested that the increasing predominance of females might be due to evolving sexual practices and a putative association with anal infections with human papilloma virus (37, 319, 320). The spatial patterns for this cancer were ranked similarly to European and Asian regions (Table 9-2), and requires more investigation.

The incidence of *thyroid* cancer increased in the 2003-07 period (6.52 per 10^5 p-y). It also had the most stable SR (0.4) in current study (even though the rate may have been influenced by different registries across time). Cook *et al* showed that the incidence was higher in

females at all ages (i.e., SR was lower in all ages) (37). The etiology for *thyroid* cancer remains poorly understood (39, 321-323). However, the high incidence at reproductive ages suggests a hormonal underpinning. There seems to be an unusual higher propensity for diagnostic investigations among female patients (37), even when there is evidence from autopsy studies that the prevalence of thyroid micro carcinoma is equivalent across sexes (324). The average annual rate is increasing for *thyroid* cancer (Appendix 35), even among males. The SRv in *thyroid* cancer was low and consistent across time in our study indicating that there might be some exposure unique only to women. The ratio of inter-intra-regional variations of SR for thyroid cancer indicates that the variation is similar in or across regions. Moran's I for thyroid cancer was ranked similarly in Europe and Asia, which also suggests that endogenous factors may be important. Again, this is speculative and requires further exploration.

Our study has provided an overview of SR variations worldwide using several analytical approaches. The data support that cancer is a complex multi-faceted disease that is heterogeneous in nature, and as such, each cancer is a different disease that might or might not share causative agents. Very much like the study using SEER data in the US population (37), most cancers investigated in this study had a high sex-ratio, suggesting a universal mechanism that increases male susceptibility to cancer. Tentative explanatory hypotheses by Cook *et al*. include sex differences in anti-oxidative capacity, metal toxicity, beliefs and behaviors, healthcare access and utilization, sex chromosome complement/ anueploidy, gene expression, hormones, immune competence, and the soma theory of aging.

Observing the consistency in SR over time, how SR changes over time, and the spatial patterns in SR provides insight into theories of cancer pathogenesis. Our study suggests that

most exposures that underpin changes in SR, need more investigation. Larger body size and high basal metabolic rates have been hypothesized to increase cancer risk in males (39, 325, 326) through greater numbers of cell proliferations (327). However, this association can also vary by sex and therefore other causes are clearly implicated. Moreover, body mass index and weight are not good markers of cell proliferation. The role of genetic or epigenetic effects needs consideration explaining SR in cancer occurrence.

Five studies have recently provided insight into SR in exploring the causes of cancers. Each used a different approach to analyze the excess risk in males and speculated on the causes of this male-female differential. Cook *et al* suggested that changing patterns in SR indicates a major role of environmental causes (37). Edgren *et al*. concluded that most of the sex differentials are enigmatic and unlikely to be explained by known environmental or purely genetic factors (39), and this point of view was further reinforced by two most recent studies (250, 328). Tomasetti and Vogelstein (281) undermined the importance of environmental and genetic factors in cancer causation by introducing a third factor (i.e., stochastic effects associated with lifetime number of stem cell divisions within each tissue of the body organs).

Wang *et. al* (329) and Rozhok *et. al* (330) labeled the study by Tomasetti and Vogelstein as a challenge to cancer prevention community. In many respects, the study was useful in inciting the scientific community to examine assumptions regarding the origins of cancer. While the statistical modeling was rigorous, the study focused on a small number of rare tumors. The analysis used incidence rates in the US only, and therefore its view of the somatic mutation theory of cancer was limited. The proposed stochastic model might seem reasonable to scientists who uphold genetics and deep sequencing of tumors as possible

causes of cancer, without regard to how mutations occur (329). The other side of picture is for those who study environmental influences on carcinogenesis and follow closely changes in cancer incidence rates over time, and across broad geographic boundaries. The conclusion that cancer rates are primarily determined by baseline stem cell numbers does not fit with most of the public health evidence and traditional thinking of many epidemiologists. An alternate and plausible explanation was provided by Cao *et al*. based on their own hypothesis that basal metabolic rate and oxidative stress levels in tissue determines the total number of cell divisions (331). Therefore, the observation by Tomasetti and Vogelstein that stochastic influences can be more of statistical correlation rather than causal.

9.2.2 Notes on intra- and inter- regional variability

Previously, Diez Roux postulated that low intra- and high inter-country variability in the incidence of the disease suggests that group-level or population level factors are important in the etiology of the disease (332). A literature search found only one study on cancer incidence trends that compared within-country and between country variability using cancer registries as surrogates for countries (79). Like our results on intra- and inter-regional variations of sex ratios, for most of the cancer types the variations in incidence were larger between-countries (inter-) than they were within-countries (intra-).

In our study, some speculation on the patterns of the causes of cancer can be made from differences in the intra- and inter-regional variations in sex ratio, based on the results of the mixed effects models (Tables 8-12 to 8-14) Cancers where intra- is somewhat similar to the inter-regional SRv ($\delta^2 \approx \tau_1^2$) (i.e., *thyroid*, *non-Hodgkin's lymphoma*, cancers of *rectum and anus, kidney, multiple myeloma, leukemia)* (Appendix 6; A 6-m and A 6-o), also show somewhat low SRm and low SRv (see Section 8.2). Therefore, not only these cancers show

consistently low SR in a cross section of five years' time but are also when analysed in terms of decades from 1983-2007. Edgren *et al* (39) has previously interpreted his similar results with the notion that an endogenous cause of the sex differential is more likely for a site where the association with sex is consistent and stable over time. However, they cautioned that such patterns may emerge also if certain environmental factors are distributed unequally between men and women worldwide.

Cancers where the intra- is less than the inter- regional variation ($\delta^2 < \tau_1^2$), also show a very high SRm and SRv from the results of analysis in section 8.2. The most notable cancers are *larynx, lung, esophagus, oral cavity* and *pharynx*, and *bladder* confirming that these are the cancers that are affected differently in males and females. Rare cancers such as of *eye* and *bone* where $\delta^2 > \tau_1^2$, may indicate a causative factor or an exposure (s) that could be specific to few particular regions of the world. The explanation of regional variations of SR for most of these cancers remains at most speculative.

9.2.3 Notes on spatial variability

There is dearth of information on spatial epidemiology of cancer incidence and mortality but a surge has been seen since 2002 (85). The purpose of carrying out spatial analysis in this study was to get an insight on how the incidence and sex ratios of different cancer types are geographically clustered and what does it mean in terms of any clues on spatial variation in causes of cancers e.g., environmental exposures, social phenomena, and other health outcomes that can enable us to apply interventions aimed at reducing the cancer burden. In this study, geographic patterns of SR are used to rank each of the 28 cancer types according to degree of clustering. Some of these similarly-ranked cancer types (in Asia and Europe) can get classified into four different types of patterns according to similarities or

dissimilarities of sex ratios in adjacent regions: 1) Spatially dependent, when more similarities are observed in adjacent geographic regions. This includes cancers of *esophagus, stomach, kidney, pancreas*, and *thyroid*; 2) Inverse spatial dependence, when there is a negative sign in Moran's index, indicating that the neighboring values are dissimilar. This includes cancers of *skin*, *multiple myeloma*, and cancer of *eye* and ranked similarly in Asia and Europe. Non-Hodgkin's lymphoma shows no specific pattern whereas lung cancer has a very different pattern in Asia and Europe. These patterns need to be explored further in terms of generating hypothesis on either exogenous or endogenous causes, as well as understanding the impact of seasonal and latitudinal effect on some of these cancers that are classified as inversely spatially dependent (333). Work of Trojano and colleagues (33) can be of great assistance in exploring the latitudinal impact on some of the cancers as and when the hypothesis permits, and by mapping spatial variations.

9.3 Tentative hypotheses and questions based on observations and trends

In this study, we analyzed sex ratios of cancer incidence by identifying types of cancers that showed consistently high or low variance of sex ratios in three different selected time periods, and through analyzing longitudinally by positing a model for changes in sex ratios over several decades. An attempt was made to see whether any clue on causation of cancers can be gained from intra- and inter-regional variations as well as through analyzing the spatial data. There is a high chance that some subtle clues in this exploratory analysis might have been missed, or we have not been able to detect other clues from the trends in sex ratios of incidences. However, one observation that is very evident is that men seems to be at higher risk of cancer apart from exceptions like *thyroid* and *gallbladder*. Apart from few attempts, there is surprisingly little research in investigations of this male-female

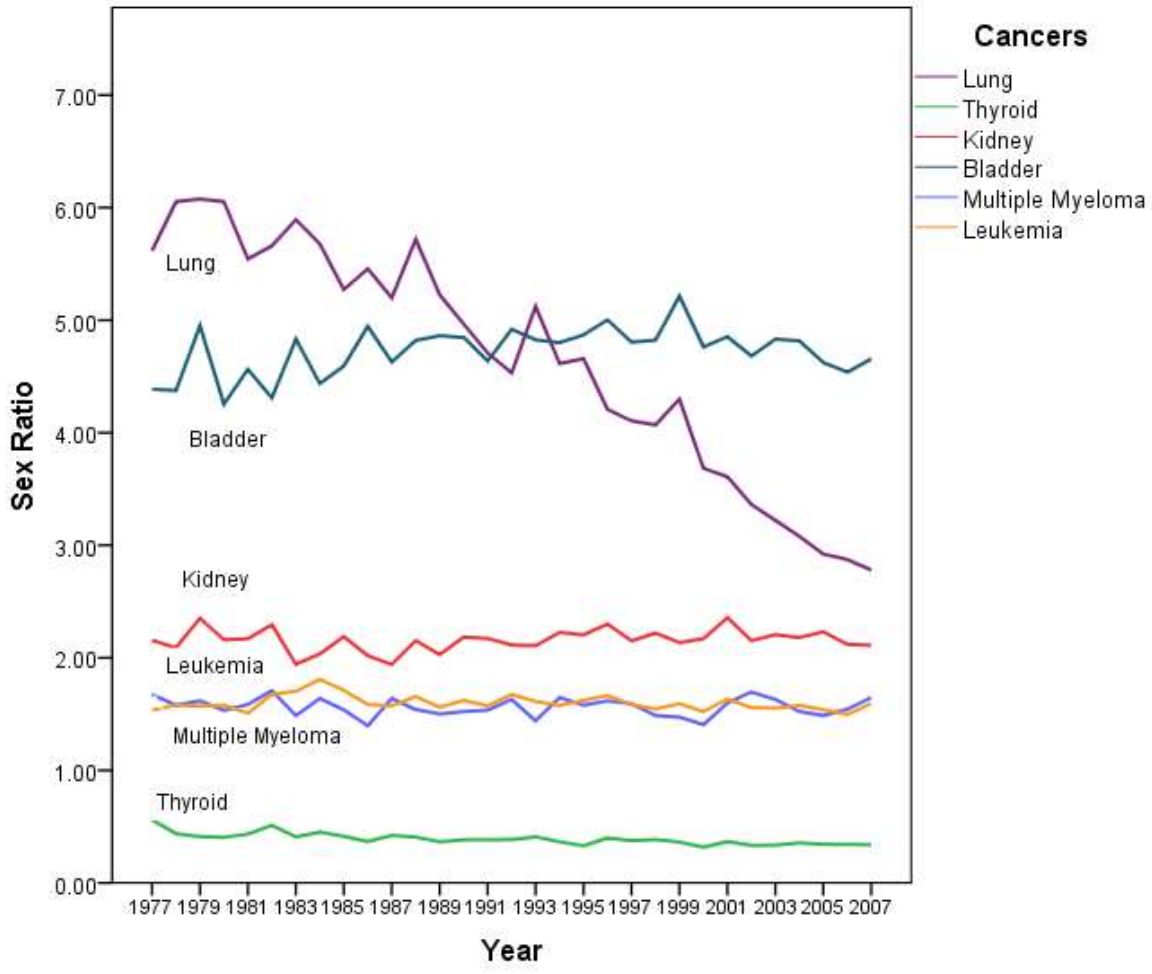
differential. A wide range of explanations are discussed in section 9.1 and 9.2 through observations in different studies as well as from this study. A prevailing view in academia for this differential is the hypothesized protection by female sex hormones that in fact is thought to account for the totality of this differential (250, 328). For some cancers, this point of view may have a standing, but for other types of cancers, this view might be erroneous.

While sex ratio magnitude and variations for some cancers are clearly correlated to variations in lifestyle factors such as in *lung* cancer, however the observed sex ratio variations for some of the other cancer types are unlikely to be explained by some known risk factors of behavioral nature e.g., tobacco smoking, alcohol drinking, and basal metabolic rate (BMI) or weight of a person. Figure 9-1 represents 76 long standing registries, and shows some of the selected cancers where causes can be completely explained such as *lung*; partly explained such as cancer of *bladder* and *kidney*; and cancers where the causes remain mostly unknown such as *leukemia*, *multiple myeloma*, and *thyroid*.

Among the high incident cancers, the SRs for *lung* cancer reflect historical exposure to tobacco smoking; as smoking habits have converged between the sexes, and so has the incidence (Figure 8-1). This pattern is also somewhat evident for other cancers for which tobacco smoking is a risk factor, including *lip*, *mouth*, and *larynx* where the incidence is decreasing in men (Figures 8-2 and 8-3). To paraphrase Cook *et al*., (37) this dramatic decrease in the SRs for cancers of the lung exemplifies the effect that a single sex-discrepant exposure can have on cancer incidence. This pattern of SR of *lung* cancer incidence in Figure 9-1 is quite palpable since the smoking prevalence between men and women have over time changed. However, what is quite unique is that smoking is not able to explain

fully, the patterns of other smoking related cancer types such as *bladder* and *kidney*, in the same manner as lung.

Figure 9-1: Sex ratios of incidence for selected cancers with stable and varying patterns by year in 76 worldwide cancer registries.



Life style behavioral factors such as smoking tobacco is the best-established risk factor for *bladder* cancer (334) and has also been strongly linked to cancers of *kidney* along with obesity and hypertension (335). It would be very natural that these factors over time will vary between men and women across many geographic regions, resulting in varying sex

ratios in incidence. Hence, this large variability in sex ratios across time and region for cancers will also be observed for risk factor such as smoking. From Figure 9-1, while it is plausible that *lung* cancer trend is due to changes in smoking prevalence, the sex ratio for another smoking associated cancer (i.e., *bladder*) has not reduced as substantially as that for *lung* cancer in recent years, suggesting the presence of other risk factors that are preserving the sex ratio for *bladder* cancer.

As established risk factors for many cancers such as *lung*, *bladder* and *kidney*, the prevalence of smoking tobacco (and obesity as another risk factor for *kidney* cancer), has changed dramatically over time. The estimated prevalence of tobacco smoking worldwide has decreased from 41% in 1980 to 31% in 2012 for men, and from 10.5% to 6% for women, with large variation in the trend in prevalence across countries (336). In the same way, the estimated prevalence of obesity increased from 3.2% to 10.8% for men, and from 6.4% to 14.9% for women, between 1975 and 2014, although the trend in prevalence varies greatly by different geographical regions (337).

In comparison to *lung* and *bladder* cancer, there is uniquely unchanging stability of the sex ratios for the cancer of *kidney* that suggests that the male-female differential (with a stable 2:1 ratio) cannot be readily explained. Figure 9-1 shows that cancer of the *kidney* is about twice as common in men as women and this ratio is maintained across many years of observations. This sex ratio stability is also evident when the analysis is stratified in younger and older age groups from IARC data on *kidney* cancer i.e., less than or more than 60 years. Hence the hypothesized protection from female sex hormones is unlikely to account for this uniquely stable sex ratio of cancer of kidney (accounting for before and after menopausal age groups).

The unwavering stability of *kidney* cancer with a sex ratio of 2:1 in both high or low incidence regions is not completely explained in the literature, especially through studies of traditionally established risk factors such as smoking and obesity. Hence our analyses have brought to surface several questions and a tentative hypothesis that can be suggested in this study, and made more specific in future as part of deductive process of scientific inquiry. *Based on our inductive process of reasoning from the observed patterns, we hypothesize that there is an unknown causative factor, likely a genetic, that contributes to the etiology of kidney cancer (although some moderate effects of smoking and a high BMI may still play a role).*

A series of rare but highly penetrant mutations have been identified, such as in the Von-Hippel Lindau (VHL gene), which has recently highlighted the importance of the VHL pathway in sporadic form of this cancer (338). Other studies have revealed insights in additional key genes for different *kidney* cancer subtypes (339-341). Through genome-wide association studies, it can be possible to detect some variants conferring susceptibility to cancers of *kidney*. Mutations of the VHL gene are associated with von Hippel-Lindau-disease, which is a hereditary cancer syndrome and predisposes to a tumor of the *eye, brain, pancreas*, and *salivary glands* in addition to *kidney*. These cancer types have also shown small variations of sex ratios in our study, and hence further exploration is needed to understand their etiology.

The analysis of intra- and inter- regional variation of sex ratio of kidney also shows that there is not much of a difference within or between regions as assessed through cancer registries in our data (Table 8-14). This pattern of similar intra- and inter- regional variability in cancer of *kidney* was also previously observed (79). Cancer of the kidney is

also a spatially dependent cancer i.e., the sex ratios are similarly clustered in geographically contiguous regions, and this similarity persists in both continents of Asia and Europe.

The observations on sex ratio of *kidney* cancer brings to light several questions that should be investigated. These could be based on a very general line of inquiry e.g., how and why major differences by ethnicity, geography, and exposure histories on smoking, alcohol, coffee drinking, obesity etc. can accentuate the complexity of etiology of this cancer in terms of male-female differentials? What possible underlying biological pathways can be postulated that can further delineate the association of smoking, obesity, and hypertension in causation of kidney cancers and their subtypes in males and females? Can a greater understanding of germline mutations (such as a gene change in human egg or sperm that then gets incorporated into the cellular DNA of the offspring) inform us about any unknown lifestyle factors and environmental exposures in the adult life? Is there a role of sex chromosomes in the etiology of kidney cancer? These and many other putative questions can become bane of studies that could explore why there is a stable sex ratio over time and geography when smoking prevalence is highly variable in men and women.

The observations in Figure 9-1 also show a very stable consistent sex ratio for cancers of *leukemia*, *multiple myeloma*, and *thyroid*, however in these cancers, role of smoking, alcohol drinking, and obesity has not been entirely conclusive. These cancers remained completely unexplained by these risk factors. For *thyroid* cancer, the sex ratio is found to be lower in younger ages unlike *kidney* cancer which is stable across all the ages. *Multiple myeloma* has also shown a consistent stable sex ratio in our study similar to cancer of *kidney*. Latitudinal effects based on stable sex ratios of multiple myeloma in our study with

small variations can be explored in a similar way as was done in a study on sex ratio in incidence of multiple sclerosis (33). Role of vitamin D deficiency has previously been hypothesized for higher incidence of multiple myeloma in some regions (342), however latitudinal effects have not been explored.

9.5 Theoretical implications of observations

Rothchild wrote his last paper on an eclectic overview of scientific inquiry and the role of induction and deduction in the methods that are practiced in the scientific community (282). According to him, making observations will inevitably generate a question, a tentative hypothesis, a hunch, and even a theory(s) about a concealed truth somewhere in the mass of findings. When the questions arise from these findings and hypotheses are suggested, the simplest way to unearth some of the answers is to accumulate more observations, perhaps more specialized ones and then theorize them.

The patterns observed in the magnitude and variations of sex ratios in conjunction with inter- and intra-regional variations and the observed spatial patterns is an exercise into an inquiry of potential cancer causes. While in most of the analyses that we did, there are opportunities of further inquiry, these observations need to be refined at a finer level. For example, the question that if there is a well-known risk factor of a cancer that predominantly varies a lot in men and women over time and place, is it possible that the same cancer will not have the same variability of sex ratio, is a hunch, that needs theorizing. A theory can only be refined, when the knowledge construction is started from the very basic level of observation e.g., as is the case of *kidney* cancer, where smoking as an established risk factor is not able to explain a stable variability of sex ratio of this cancer. In this section, some observations are presented at very basic level that requires

collection of more observations, and hence provide fodder for generalization. Generalization, which is another term used for theorizing, is the last step in the process of inductive inference and the first step in the deductive reasoning. This also means how observations are interpreted and becomes a point of view.

Lutz and Fekete used the ratio of low- and high-incidence registries to speculate on the role of exogenous (avoidable) and endogenous (non-avoidable) factors in carcinogenesis for numerous types of cancers (253). In simple terms, the patterns of male and female incidence likely arise due to differences in exposures to environmental carcinogens, behavioral factors, to expression of different hormonal profiles, or different patterns of tissue growth, damage, or repair, as well as genetics and heritable susceptibility. In the following paragraphs, we present some generalization as part of the induction process itself that could become a starting point for deductive inferences where more specific hypotheses can be presented and tested in future.

In regions or countries where the prevalence of exogenous exposures differs in males and females, there is high variation in the SR of cancer incidence (provided that the exposure is higher in males). We speculate behavioral and environmental causes are more likely than endogenous causes for cancers in which both the SRm and SRv are high.

Most of the cancers have higher risks in men. Cancer types for which the dominant etiology is alcohol show the widest variations in SR with most wine-drinking countries at the top (in all three selected time periods). For cancers of the *esophagus, larynx, pharynx, lip, mouth and tongue*, consistently high SRm and SRv (Tables 8-6 to 8-11) were observed in France, Spain and registries from Adriatic and Baltic regions (Tables 8-3 to 8-5; and also Appendix 4). *Liver* cancer, which is now a major public health problem in developing

countries due to the high prevalence of hepatitis infections, also had higher SRm in registries of developed countries, and more specifically in several regions in France (Appendix 4). Although SRv in liver cancer is not as high as in other alcohol-related cancers, the patterns indicate a prominent etiological role of alcohol consumption in European countries.

In *lung* cancer where the dominant cause is smoking (and men and women are differentially exposed), the SRm was highest in France and Spain (1973-77 and 1988-92), and in Turkey, Belarus and Tunisia in 2003-07 (Appendix 4). The lowest SR of *lung* cancer (i.e., relative female predominance) were observed in the Northwest Territories in Canada, Cuba, Iceland, and regions in India. *Bladder* cancer, another smoking-related cancer, showed higher magnitude and variation in Mediterranean regions.

In regions or countries where the causes of cancers are similar in males and females, SRv of cancer incidence is low. For cancers where both SRm and SRv are low, endogenous factors are likely predominant causes. However, exogenous factors can also be compatible with the observation.

The consistently low magnitude of SR and its variation in *kidney, leukemia, multiple myeloma and thyroid* cancer (Tables 8-6 to 8-11) support the hypothesis that for these cancer sites, endogenous factors are more likely. Environmental factors such as sunlight, radiation, and diet may be dominant factors for *skin*, *leukemia*, *non-Hodgkin's lymphoma*, and *rectal* cancers. Narrow SR ranges for digestive tract cancers such as *colon, small intestine, rectum*, and *stomach* and their stability (strongly stable in colon, rectum and small intestine; moderately strong in stomach) over time, is consistent with the hypothesis

that nutritional factors have a role in these cancers, although the role of *helicobacter* infections cannot be downplayed especially for stomach cancers.

For cancer of *kidney*, we inferred that there is a possibility of genetic component that plays a role in consistently stable sex ratio across time, region, as well as different age groups. One contributor to cancer incidence is indeed genetic variation, and 10% of cancers have this heritable component (343, 344). When hereditary genetic factors are identified, the way in which they contribute to differences in cancer incidence, however remains obscure. Tomasetti and Vogelstein (281) raised the question that if environmental and genetic factors cannot fully explain differences in cancer risk in different body organs, how can these differences be explained. They suggested a unique theory that stochastic (i.e., random) influences associated with lifetime number of stem cell divisions in each body organ, can in fact be *major* contributors to cancer incidence, unlike non-stochastic factors in carcinogenesis such as tobacco, alcohol, viruses, obesity, sun light exposure, dietary factors and genetic factors. However, they also made a provocative statement that most cancers are due to biological *bad-luck* and cannot be attributed to specific preventable causes. This conjecture was reported widely in the media (345-348) and was met with uniformly harsh critique in the scientific community (349-353). Though controversial, their work presents a unique theoretical perspective that could be further explored with a well-informed specific hypothesis as part of deductive inference in future.

9.6 Differential case-ascertainment: potential of gender-bias

Gender-bias is well-investigated in the field of cardiovascular diseases and are associated with behavior and life experiences (354). For example in developed countries, this includes differences in the use of prescription of drugs, health insurance reimbursement,

referral or acceptance for a particular therapy whereas in developing countries this could be as basic as access to health care diagnostic facilities and treatment (355). Gender-bias issues are reflected in about 10,000 scientific articles on wide range of diseases that includes diseases of pulmonary and autoimmune systems, hematology, neurology and other clinical sciences, yet there are less than 300 studies that deal with this issue in the field of oncology (355).

With regard to possible gender-bias in cancer registries, to date, there is only one study that has highlighted underreporting by sex and investigated SR in cancer registrations (182). The study focused on the SR of childhood cancers (in less than 14 years aged boys and girls) for the period 1980-1989 from the International database on Incidence of Childhood Cancer (356). They analyzed registration ratios (boys:girls) with gross domestic product (GDP) and infant mortality rates in 53 developed and developing countries. The investigators provided several possible explanations for higher observed SR in their study: that they may reflect the true differences in the populations at risk, or that it could be due to differential early survival in boys and girls. A more likely explanation for higher SR that the authors provided was for countries like Western India and Bangladesh where sick boys are more likely to reach specialist medical centers than girls. The investigators inferred that in some of the poorest countries, the elevated SR reflects socio-economic level more than the underlying nature and etiology of the cancers. However, they also believed that GDP is not a good indicator of social and economic development of a country since it provides unreliable estimates of income.

With regard to our study on SRs, despite the fact that the data provided by individual registries undergoes rigorous quality checks and scrutiny for accuracy and completeness,

by IACR staff, before their final inclusion in the CI-5 (357); we acknowledge that there could still be a possibility of differential case ascertainment by sex. For this reason, we examined few cancers for whom smoking was risk factor in registries located in countries with low and high gender inequality.

Table 8-17 (a and b) shows results of SR for cancers of lung, bladder, esophagus, and larynx from cancer registries of selected countries (regional registries included China-Beijing and India-Mumbai). In more developed countries with very low gender inequality such as Sweden, Denmark, Norway, Iceland, and New Zealand where the prevalence of smoking in men and women is almost the same, the SR of lung cancer in these registries is lower than the overall worldwide SR of 3.3 (Table 8-2 and 8-5). For these registries, it seems that the ascertainment of cancer cases would be more complete. Although the prevalence of smoking is also somewhat similar in countries such as Finland and Belgium, the observed SR for lung cancer is on the higher side (3.2 and 3.4). In Czech Republic, more males are smokers and the SR of lung cancer is relatively high as well as for other cancers. The most unexpected deviance that we see is from Spain where the sex ratios for the selected cancers were high when in fact the smoking prevalence is similar in Spain. This probably reflects not only quality issues but there might be gender bias in Spanish cancer registries. It could also be possible that there might be some outlier registries where females might not have the same access to health care services and hence the inflated sex ratios for the whole of Spain that is ranked eleventh out of 160 countries.

Among the cancer registries with moderate level of gender inequality (China, Bahrain, Kuwait, and Ecuador), only Ecuador that shows similar patterns of smoking in males and females, has a SR of cancers that reflects the rates in lowest gender inequality countries.

While China resembles Czech Republic in terms of male-female differences in smoking prevalence, the SR of selected cancers are very low in China. The most interesting patterns are seen in Bahrain and Kuwait where the SRs of cancers are lower than Finland, Belgium and Austria even though the prevalence of smoking in males is very high in both countries. Among the countries that have high gender inequality index, only the registry in Qatar shows the faint possibility of higher gender-bias in cancer registration. The SR of lung cancer is very high in Qatar compared to other registries. It is also interesting to note that bladder and laryngeal cancers show relatively similar patterns in both high and low gender inequality countries while esophageal cancers show an intriguing pattern in high gender inequality registries where seemingly more women than men are affected compared to low gender inequality countries.

An attempt was also made to examine whether the changes in sex ratios over time are significantly different between low and high gender inequality countries. The trajectories of change in each of the four selected cancers did not show any difference. This can at some level speak of quality of long term cancer registries from CI5plus that can be deemed good compared to individual cross sections of time as defined by ten volumes of CI5.

9.7 Strengths and weaknesses

Unlike previous studies in which only the magnitude of sex ratio was reported to understand the gaps in understanding male-female differentials (37), we systematically examined the variability of sex ratios to make inferences from the explainable and unexplainable trends of male and female incidence rates. The analysis was facilitated by the availability of large number of cancer registries from different continents. To

understand the possibility of bias that can be created through differential access to health care services by men and women over time in different regions, we examined four smoking-associated cancers through a multidimensional indicator of gender inequality for the first time in the context of cancer registration. Although this bias assessment was not made through technically advanced sophisticated methods that could have measured it, nonetheless, the analysis provided us some clues that bias could be present regardless of the registry being in low and high gender inequality countries. Although presumably this bias could be more prominent in data from developing countries, we did not find any such indication.

The lack of information on individual level data on risk factors such as smoking, alcohol drinking, exposure to occupational carcinogens, socioeconomic status, and high BMI in our study means that we cannot directly assess their effects on sex ratios of different cancer types, and hence an important weakness of this study. However, by drawing attention to varying distribution of these confounding factors, we have facilitated investigators in future to find yet some of the overlooked causes that can explain the high or low disparities in males and females by making linkages to relevant medical informatics databases.

Many cancers have multiple anatomical subsites and/or multiple histological types. It is possible that some subtypes have distinct patterns reflecting distinct etiologies. By classifying cancers by broad anatomy only, the present analysis might have obscured etiological heterogeneity by site of tumor and histology e.g., proximal versus distal colon and squamous cell carcinoma versus adenocarcinoma of esophagus. Furthermore, our assessment of endogenous and exogenous causes of cancers in terms of theorizing them

was very broad. Our categorization of causes into endogenous versus exogenous was not intended to be definitive, but rather a very general preface to illustrate the potential for further inquiry as part of deductive inference.

This thesis could have been restricted to a small number of cancers, which would have permitted developing and exploring etiological hypotheses in more depth. However, variations in cancer risk by geographic region and changes over time are best demonstrated by the analyses conducted. The analysis was the most convenient strategy to address large number of cancers and to speculate on the relative importance of endogenous and exogenous contributors. The cost-effectiveness in retrieving a large volume of information was equivalent regardless of the number of cancers investigated.

This study used different number of cancer registries in three different time periods and hence it could be possible that sex ratios observed might be biased because of newer registries added up in the latter time period. However, the analysis on the long standing registries might be able to dilute some of the limitations that are observed in the analysis of cross section of three time. Nevertheless, the three-time periods should maintain equal number of cancer registries for making valid inferences. In the long-standing registries from CI5Plus, our categorization of intra- and inter-regional variability is very broad and that the definition and analyses should be revisited in future. Same will be true with the spatial analyses from a modified version of Moran's Index. However, the index provides some inclination that some of the cancers can have latitudinal effects, but the analyses need further improvement through mapping the geography.

Both incidence rate differences and rate ratios are useful measures of sex differences but provide different information. Choosing a specific parameter reflects different

perspectives about which information is important to capture. For example, the incidence rate in males and females of a particular cancer might decrease progressively over time, which is easily identifiable by an absolute difference in incidence. However, use of a relative scale (i.e., sex ratio of incidence rates) in the same example could yield an increasing trend. Therefore, the data can produce contradictory interpretations depending on which scale is used. In health-related research, the relative scale is used more often. Yet it is equally important to know the interpretation on an absolute scale (i.e., incidence rate difference). Currently there is no consensus on the most appropriate scale to use (358) . Recording differences on relative scale in SR is more likely to yield stable and reliable indicators considering possibility of imperfect ascertainment and presence of other artifacts in registries worldwide. Moreover, sex differences on an absolute scale can be confounded by the incidence in the area.

9.8 Conclusion

In conclusion, this exploratory analysis on sex ratio variations on cancer incidence provided an eclectic and wide-ranging overview on temporal and geographic trends of incidence rates. The inductive process of inferences, from observations to tentative hypotheses to theoretical generalizations provides further encouragement in utilizing sex-ratio as a useful method to further reappraise the trends in a deductive manner, and to generate specific hypotheses. The sex ratio variations, large or small, for different types of cancers is a fodder to carry out well-designed population based studies on individual cancer types, and then test the specific hypotheses. Hence, the demonstration of an uneven distribution of sex ratios among populations from different geographical areas, derived from large international body of cancer registries, adds useful information for planning

future studies that can be aimed to explore the paths of etiological hypotheses on cancers, primarily where male and female discrepancies (i.e., higher risks in men than women) cannot be fully explained by known risk factors such as smoking, alcohol, and even obesity. This study identified few cancer types such as *multiple myeloma, leukemia*, and *thyroid*, but most notably cancer of the *kidney* where the suspected and commonly acknowledged risk factor cannot fully explain the observed sex ratio variation that is consistent across time, geography and even age group. The well-established risk factors such as smoking and obesity whose prevalence varies in both sexes worldwide, is not able to decipher the curiously stable sex ratios in cancer of *kidney*. This observation has made us to tentatively hypothesize that some endogenous factor such as a gene or gene variant might be able to explain the stable sex ratio of this cancer. Another cancer type, *multiple myeloma* is also consistently stable across time and place and the role of vitamin D has previously been postulated. In this study, there is a subtle indication that the latitudinal effects can be tentatively hypothesized in multiple myeloma, but that certainly needs to be further advanced in the form of specific hypothesis and analyzed accordingly. In such a case, a well-designed spatial analytic study can be used in both inductive and deductive manner to further discern the trends that have been observed in this study.

The findings from this study indicates that for many cancer types, the sex ratio is more than two, across many years of observations. Observations of varying sex discrepancy of more than two-fold for a long period of time for some cancers does indicate a predominant role of exogenous factors. However, caution is still needed, when we attribute these factors to male excess in cancer risk. This male excess in cancer risk is well known, yet surprisingly it is still poorly studied, because in part, as Edgren (39) and Radkiewicz (250)

points out, many researchers assume that this excess is due to known causes. Therefore, region-specific estimates of the prevalence of known exposures also need to be examined to see whether these variations between men and women are correlated to changes in sex ratios of cancer incidence. This strategy could perhaps unveil the role of some of the unknown or yet undiscovered causes of cancers.

In our study, fluctuations of SRs in incidence across countries with low and high gender inequality, according to smoking prevalence, offer some clue of gender-bias that should be explored further. Genetic and biological factors could be the reasons for higher incidence in males (359, 360). The international variation and temporal changes in the magnitude of gender-bias should be studied in depth because this differential is also associated with environmental and occupational exposures, and psycho-social factors. More importantly, in the current study, gender-dependent disparities in health care referral patterns might exists in registries of countries with low gender inequality, however as seen in Spain, registries could be subject to this bias in high resource countries.

The study points towards important gaps in our understanding of causes of cancer risk in populations. From a prevention perspective, understanding the contribution of known or unknown causes along with the potential gender-disparities in cancer reporting is critical to public health practitioners and policy makers. Our study conceals some role of gender disparities in cancer registries, that needs to be further studied. The combination of 'epidemiological top-down or inductive approach', along with the 'laboratory's bottom-up or deductive approach' can yield scientific and medical advantages thereby enhancing prevention and treatments of cancer.

9.9 Contribution to knowledge

The current study adds to a very small body of existing literature that have utilized sex-ratios to understand variations in incidence trends of cancers. A large number of population based cancer registries were utilized to observe wide ranging trends across time and geography. The study confirms like previous work that risk of cancer is disproportionately higher in men for majority of cancer types. A systematic evaluation of variability of sex ratios has uncovered sex specific incidence rates for some of the cancers for which there is no likely explanation through well established risk factors. These cancers include *kidney*, *multiple myeloma*, *leukemia*, and *thyroid* cancer. The study has identified cancer of *kidney* as a unique case where the sex ratio variability is remarkably stable with an approximate ratio of 2:1 across many years of observations and regions, and the age groups. The stability also seems to be consistent regardless of high and low incidence regions for this cancer. This finding has implications since the incidence of *kidney* cancer is rising in many parts of the world. The study has demonstrated the potential of sex ratios as a robust method that can unveil concealed patterns, to assist in proposing tentative hypotheses, and that can have both public health and scientific implications. Public health questions can be formulated to ask why the incidence of one particular cancer is on increase in specific regions versus the stability of sex ratio of same cancer in other regions? What is the impact of varying environmental causes on these trends? How much role can non-environmental causes play in the etiology of cancers when it cannot be explained through well recognized environmental causes? These questions not only can contribute to epidemiology of disease but can also provide a blueprint to investigators interested in studying underlying biology of specific cancer.

References

1. Bray F, Jemal A, Grey N, Ferlay J, Forman D. Global cancer transitions according to the Human Development Index (2008-2030): a population-based study. Lancet Oncol. 2012;13(8):790-801.

2. WHO. Global Status Report on Non-Communicable Diseases 2014. http://www.who.int/nmh/publications/ncd-status-report-2014/en/ 2014.

3. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015;136(5):E359-86.

4. Bray F, Soerjomataram I. The Changing Global Burden of Cancer: Transitions in Human Development and Implications for Cancer Prevention and Control. In: Gelband H, Jha P, Sankaranarayanan R, Horton S, editors. Cancer: Disease Control Priorities, Third Edition (Volume 3). Washington (DC)2015.

5. Izquierdo JN, Schoenbach VJ. The potential and limitations of data from population-based state cancer registries. Am J Public Health. 2000;90(5):695-8.

6. Bray F. Temporal studies of cancer occurrence and applications of age-period-cohort method to trends in Europe. London: University of London; 2005.

7. Bray F, Dos Santos Silva I, Moller H, Weiderpass E. Endometrial cancer incidence trends in Europe: underlying determinants and prospects for prevention. Cancer Epidemiol Biomarkers Prev. 2005;14(5):1132-42.

8. Bray F, Loos AH, Oostindier M, Weiderpass E. Geographic and temporal variations in cancer of the corpus uteri: incidence and mortality in pre- and postmenopausal women in Europe. Int J Cancer. 2005;117(1):123-31.

9. Loos AH, Bray F, McCarron P, Weiderpass E, Hakama M, Parkin DM. Sheep and goats: separating cervix and corpus uteri from imprecisely coded uterine cancer deaths, for studies of geographical and temporal variations in mortality. Eur J Cancer. 2004;40(18):2794-803.

10. Bray F. The Evolving Scale and Profile of Cancer Worldwide: Much Ado About Everything. Cancer Epidemiol Biomarkers Prev. 2016;25(1):3-5.

11. Esteve J. International study of time trends. Some methodological considerations. Ann N Y Acad Sci. 1990;609:77-84; discussion -6.

12. Parkin DM, Fernandez LM. Use of statistics to assess the global burden of breast cancer. Breast J. 2006;12 Suppl 1:S70-80.

13. Bray F, Kohler B, Ferlay J. Data comparibility and quality. Cancer Incidence in Five Continents. IARC Scientific Publication No. 165. Lyon: International Agency for Research on Cancer; 2014.

14. Gardner MJ, Osmond C. Interpretation of disease time trends: is cancer on the increase? A simple cohort technique and its relationship to more advanced models. J Epidemiol Community Health. 1983;37(4):274-8.

15. Gardner MJ, Osmond C. Interpretation of time trends in disease rates in the presence of generation effects. Stat Med. 1984;3(2):113-30.

16. Bray F, Parkin DM. Evaluation of data quality in the cancer registry: principles and methods. Part I: comparability, validity and timeliness. Eur J Cancer. 2009;45(5):747-55.

17. Parkin DM, Bray F. Evaluation of data quality in the cancer registry: principles and methods Part II. Completeness. Eur J Cancer. 2009;45(5):756-64.

18. Moore MA, Shin HR, Curado MP, Sobue T. Establishment of an Asian Cancer Registry Network - problems and perspectives. Asian Pac J Cancer Prev. 2008;9(4):815-32.

19. Parkin DM, Bray FI, Devesa SS. Cancer burden in the year 2000. The global picture. Eur J Cancer. 2001;37 Suppl 8:S4-66.

20. Li GL, Chen WQ. Representativeness of population-based cancer registration in China--comparison of urban and rural areas. Asian Pac J Cancer Prev. 2009;10(4):559-64.

21. Jensen OM, Storm HH. Cancer registration: principles and methods. Reporting of results. IARC Sci Publ. 1991(95):108-25.

22. Jensen OM, Storm HH. Purposes and uses of cancer registration. IARC Sci Publ. 1991(95):7-21.

23. Doll RP, P. Waterhouse, JAH. . Cancer Incidence in Five Continents, Volume 1. Union Internationale Contre le Cancer, Geneva. 1966.

24. Ferlay J, Bray F, Steliarova-Foucher E, Forman D. Cancer Incidence in Five Continents, CI5plus. IARC CancerBase No. 9. Lyon: International Agency for Research on Cancer; 2014. Available from: http://ci5.iarc.fr, accessed [June 20, 2016]. 2014.

25. Forman D, Bray, F., Brewster, D.H., Gombe Mbalawa, C., Kohler, B., Piñeros, M., Steliarova-Foucher, E., Swaminathan, R., and Ferlay, J. Cancer Incidence in Five Continents, Vol. X. Lyon: International Agency for Research on Cancer.; 2014.

26. Doll R, Payne P, Waterhouse J. Cancer incidence in five continents: a technical report. Berlin: Springer-Verlag (for UICC) 1966.

27. Parkin DM. The evolution of the population-based cancer registry. Nat Rev Cancer. 2006;6(8):603-12.

28. Parkin DM, Chen VW, Ferlay J, Galceran J, Storm H, Whelan S. Comparability and quality control in cancer registration. (IARC Technical Report No. 19). Lyon: IARC (WHO) and IACR. 1994.

29. Saxen E. Trends: facts or Fallacy. Trends in Cancer Incidence: Causea and Practical Implications. Oslo: The International Union Against Cancer and The Norwegian Cancer Society; 1982. p. 5-16. 1982.

30. Muir CS, Fraumeni JF, Jr., Doll R. The interpretation of time trends. Cancer Surv. 1994;19-20:5-21.

31. Swerdlow AJ, Douglas AJ, Vaughan Hudson G, Vaughan Hudson B. Completeness of cancer registration in England and Wales: an assessment based on 2,145 patients with Hodgkin's disease independently registered by the British National Lymphoma Investigation. Br J Cancer. 1993;67(2):326-9.

32. Palacios N, Alonso A, Bronnum-Hansen H, Ascherio A. Smoking and increased risk of multiple sclerosis: parallel trends in the sex ratio reinforce the evidence. Ann Epidemiol. 2011;21(7):536-42.

33. Trojano M, Lucchese G, Graziano G, Taylor BV, Simpson S, Jr., Lepore V, et al. Geographical variations in sex ratio trends over time in multiple sclerosis. PLoS One. 2012;7(10):e48078.

34. Zhao J, Booth H, Dear K, Tu EJ. Cardiovascular mortality sex differentials in selected East Asian and Western populations. J Epidemiol Community Health. 2016;70(10):983-9.

35. Lawlor DA, Ebrahim S, Davey Smith G. Sex matters: secular and geographical trends in sex differences in coronary heart disease mortality. BMJ. 2001;323(7312):541-5.

36. Orton SM, Herrera BM, Yee IM, Valdar W, Ramagopalan SV, Sadovnick AD, et al. Sex ratio of multiple sclerosis in Canada: a longitudinal study. Lancet Neurol. 2006;5(11):932-6.

37. Cook MB, Dawsey SM, Freedman ND, Inskip PD, Wichner SM, Quraishi SM, et al. Sex disparities in cancer incidence by period and age. Cancer Epidemiol Biomarkers Prev. 2009;18(4):1174-82.

38. Cook MB, McGlynn KA, Devesa SS, Freedman ND, Anderson WF. Sex disparities in cancer mortality and survival. Cancer Epidemiol Biomarkers Prev. 2011;20(8):1629-37.

39. Edgren G, Liang L, Adami HO, Chang ET. Enigmatic sex disparities in cancer incidence. Eur J Epidemiol. 2012;27(3):187-96.

40. Bray F, Ferlay J, Laversanne M, Brewster DH, Gombe Mbalawa C, Kohler B, et al. Cancer Incidence in Five Continents: Inclusion criteria, highlights from Volume X and the global status of cancer registration. Int J Cancer. 2015;137(9):2060-71.

41. Shin HR, Masuyer E, Ferlay J, Curado MP, Asian Contributors to CI. Cancer in Asia - Incidence rates based on data in cancer incidence in five continents IX (1998-2002). Asian Pac J Cancer Prev. 2010;11 Suppl 2:11-6.

42. Sherman RL, Boscoe FP, O'Brien DK, George JT, Henry KA, Soloway LE, et al. Misclassification of sex in central cancer registries. J Registry Manag. 2014;41(3):120-4.

43. Din NU, Ukoumunne OC, Rubin G, Hamilton W, Carter B, Stapley S, et al. Age and Gender Variations in Cancer Diagnostic Intervals in 15 Cancers: Analysis of Data from the UK Clinical Practice Research Datalink. PLoS One. 2015;10(5):e0127717.

44. Lyratzopoulos G, Abel GA, McPhail S, Neal RD, Rubin GP. Gender inequalities in the promptness of diagnosis of bladder and renal cancer after symptomatic presentation: evidence from secondary analysis of an English primary care audit survey. BMJ Open. 2013;3(6).

45. Permanyer I. The Measurement of Multidimensional Gender Inequality: Continuing the Debate. Soc Indic Res. 2010;95:181-98.

46. Permanyer I. A Critical Assessment of the UNDP's Gender Inequality Index. Feminist Economist. 2013;19(2):1-32.

47. Hassanzadeh JM, N; Esmailnasab, N; Rezaeian, S; Bagheri, P: Armanmehr, V. The Correlation between Gender Inequalities and Their Health Related Factors in World Countries: A Global Cross-Sectional Study. Epidemiology Research Int. 2014:1-8.

48. Singh GK, Azuine RE, Siahpush M. Global Inequalities in Cervical Cancer Incidence and Mortality are Linked to Deprivation, Low Socioeconomic Status, and Human Development. Int J MCH AIDS. 2012;1(1):17-30.

49. Gender Inequality Index. United Nations Devlopment Programme: Human Development Reports. Accessed March 31, 2017. [Available from: http://hdr.undp.org/en/content/gender-inequality-index-gii.

50. Mason KO. Explaining fertility transitions. Demography. 1997;34(4):443-54.

51. Smyth I. Gender analysis of family planning: Beyond the feminist vs. population control debate. Feminist Economist. 1996;2(2):63-86.

52. Klasen S. Does gender inequality reduce growth and development? Evidence from cross-country regressions. Policy research report on gender and development, Working paper series, No 7, The World Bank. 1999.

53. Dollar D, Gatti R. Gender inequality, income and growth: Are good times for women? Mimeographed, Washington DC: The World Bank. 1999.

54. Grant WB. An ecological study of cancer mortality rates in the United States with respect to solar ultraviolet-B doses, smoking, alcohol consumption and urban/rural residence. Dermatoendocrinol. 2010;2(2):68-76.

55. Grant WB. An ecological study of cancer incidence and mortality rates in France with respect to latitude, an index for vitamin D production. Dermatoendocrinol. 2010;2(2):62-7.

56. Grant WB. A multicountry ecological study of risk-modifying factors for prostate cancer: apolipoprotein E epsilon4 as a risk factor and cereals as a risk reduction factor. Anticancer Res. 2010;30(1):189-99.

57. Grant WB. Ecological studies of the UVB-vitamin D-cancer hypothesis. Anticancer Res. 2012;32(1):223-36.

58. Kaiser BB, K. What predicts breast cancer rates? Testing hypotheses of the demographic and nutrition transitions. J Pop Research. 2013;30(1):67-85.

59. Mohr SB, Gorham ED, Garland CF, Grant WB, Garland FC. Low ultraviolet B and increased risk of brain cancer: an ecological study of 175 countries. Neuroepidemiology. 2010;35(4):281-90.

60. Kamangar F, Dores GM, Anderson WF. Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. J Clin Oncol. 2006;24(14):2137-50.

61. Garland CF, Garland FC. Do sunlight and vitamin D reduce the likelihood of colon cancer? Int J Epidemiol. 1980;9(3):227-31.

62. Garland FC, Garland CF, Gorham ED, Young JF. Geographic variation in breast cancer mortality in the United States: a hypothesis involving exposure to solar radiation. Prev Med. 1990;19(6):614-22.

63. Grant WB, Garland CF. The association of solar ultraviolet B (UVB) with reducing risk of cancer: multifactorial ecologic analysis of geographic variation in age-adjusted cancer mortality rates. Anticancer Res. 2006;26(4A):2687-99.

64. Hartge P, Devesa SS, Grauman D, Fears TR, Fraumeni JF, Jr. Non-Hodgkin's lymphoma and sunlight. J Natl Cancer Inst. 1996;88(5):298-300.

65. Lefkowitz ES, Garland CF. Sunlight, vitamin D, and ovarian cancer mortality rates in US women. Int J Epidemiol. 1994;23(6):1133-6.

66. Mohr SB. A brief history of vitamin d and cancer prevention. Ann Epidemiol. 2009;19(2):79-83.

67. Schwartz GG, Hulka BS. Is vitamin D deficiency a risk factor for prostate cancer? (Hypothesis). Anticancer Res. 1990;10(5A):1307-11.

68. Grant WB. The effect of solar UVB doses and vitamin D production, skin cancer action spectra, and smoking in explaining links between skin cancers and solid tumours. Eur J Cancer. 2008;44(1):12-5.

69. Kjaer SK, de Villiers EM, Haugaard BJ, Christensen RB, Teisen C, Moller KA, et al. Human papillomavirus, herpes simplex virus and cervical cancer incidence in Greenland and Denmark. A population-based cross-sectional study. Int J Cancer. 1988;41(4):518-24.

70. Higginson J, Muir CS. Environmental carcinogenesis: misconceptions and limitations to cancer control. J Natl Cancer Inst. 1979;63(6):1291-8.

71. Higginson J, Oettle AG. Cancer incidence in the Bantu and "Cape Colored" races of South Africa: report of a cancer survey in the Transvaal (1953-55). J Natl Cancer Inst. 1960;24:589-671.

72. Doll R, Peto R. The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. J Natl Cancer Inst. 1981;66(6):1191-308.

73. Parkin DM. International variation. Oncogene. 2004;23(38):6329-40.

74. Armstrong BK. The role of the cancer registry in cancer control. Cancer Causes Control. 1992;3(6):569-79.

75. Wagner G. History of cancer registration. IARC Sci Publ. 1991(95):3-6.

76. Saxen E. Trends in cancer incidence: causes and practical implications. . Magnus K, editor. Washington, D.C: Hemisphere Publishing Corporation; 1982. 446 p.

77. Luoto R, Raitanen J, Pukkala E, Anttila A. Effect of hysterectomy on incidence trends of endometrial and cervical cancer in Finland 1953-2010. Br J Cancer. 2004;90(9):1756-9.

78. Wang YC, Wei LJ, Liu JT, Li SX, Wang QS. Comparison of Cancer Incidence between China and the USA. Cancer Biol Med. 2012;9(2):128-32.

79. Stang A, Kowall B, Rusner C, Trabert B, Bray F, Schuz J, et al. A novel method for identifying settings for well-motivated ecologic studies of cancer. Int J Cancer. 2016;138(8):1887-93.

80. Crocetti E, Giusti F, Martos C, Randi G, Dyba T, Bettio M. Variability of cancer risk within an area: time to complement the incidence rate. Eur J Cancer Prev. 2017;26(5):442-6.

81. Kelsall J, Wakefield J. Modelling spatial variation in disease risk: a geostatistical approach. J Am Stats Asoc. 2002;97(459):692-01.

82. Rogerson PA. Spatial Patterns: In Spatial Methods for Geography: Sage Publications; 3 edition.; 2010.

83. Boscoe FP, Ward MH, Reynolds P. Current practices in spatial analysis of cancer data: data characteristics and data sources for geographic studies of cancer. Int J Health Geogr. 2004;3(1):28.

84. Dockerty JD, Sharples KJ, Borman B. An assessment of spatial clustering of leukaemias and lymphomas among young people in New Zealand. J Epidemiol Community Health. 1999;53(3):154-8.

85. Roquette R, Painho M, Nunes B. Spatial epidemiology of cancer: a review of data sources, methods and risk factors. Geospat Health. 2017;12(1):504.

86. Hall A, Ahonen-Rainio P, Virrantaus K. Knowledge and Reasoning in Spatial Analysis. John Wiley & Sons Ltd Transactions in GIS, volume 18, issue 3, pages 464-476 ISSN 1467-9671 DOI 101111/tgis 12049. 2014.

87. Boyle P. Relative value of incidence and mortality data in cancer research. Recent Results Cancer Res. 1989;114:41-63.

88. Percy C, Ries LG, Van Holten VD. The accuracy of liver cancer as the underlying cause of death on death certificates. Public Health Rep. 1990;105(4):361-7.

89. Percy C, Stanek E, 3rd, Gloeckler L. Accuracy of cancer death certificates and its effect on cancer mortality statistics. Am J Public Health. 1981;71(3):242-50.

90. Percy CL, Miller BA, Gloeckler Ries LA. Effect of changes in cancer classification and the accuracy of cancer death certificates on trends in cancer mortality. Ann N Y Acad Sci. 1990;609:87-97; discussion -9.

91. Percy C, Dolman A. Comparison of the coding of death certificates related to cancer in seven countries. Public Health Rep. 1978;93(4):335-50.

92. Grulich AE, Swerdlow AJ, dos Santos Silva I, Beral V. Is the apparent rise in cancer mortality in the elderly real? Analysis of changes in certification and coding of cause of death in England and Wales, 1970-1990. Int J Cancer. 1995;63(2):164-8.

93. Berrino F, Gatta G, Sant M, Capocaccia R. The EUROCARE study of survival of cancer patients in Europe: aims, current status, strengths and weaknesses. Eur J Cancer. 2001;37(6):673-7.

94. Pisani P, Bray F, Parkin DM. Estimates of the world-wide prevalence of cancer for 25 sites in the adult population. Int J Cancer. 2002;97(1):72-81.

95. Sant M, Aareleid T, Berrino F, Bielska Lasota M, Carli PM, Faivre J, et al. EUROCARE-3: survival of cancer patients diagnosed 1990-94--results and commentary. Ann Oncol. 2003;14 Suppl 5:v61-118.

96. Sant M, Allemani C, Santaquilani M, Knijn A, Marchesi F, Capocaccia R, et al. EUROCARE-4. Survival of cancer patients diagnosed in 1995-1999. Results and commentary. Eur J Cancer. 2009;45(6):931-91.

97. Irwig L, Armstrong B. EUROCARE-2: relevance for assessment of quality of cancer services? Lancet. 2000;355(9202):427-8.

98. Muir CS, Percy C. Cancer registration: principles and methods. Classification and coding of neoplasms. IARC Sci Publ. 1991(95):64-81.

99. Clemmesen J, Nielsen A. Morbidity versus mortality. J Natl Cancer Inst. 1982;69(3):549-50.

100. Devesa SS, Pollack ES, Young JL, Jr. Assessing the validity of observed cancer incidence trends. Am J Epidemiol. 1984;119(2):274-91.

101. Pastides H. The descriptive epidemiology of cancer. Fundamentals of cancer epidemiology. Sudburry, Massachussets: Jones and Bartlett Publishers; 2008. p. 531.

102. Cleveland WS, McGill R. Graphical perception and graphical methods for analyzing scientific data. Science. 1985;229(4716):828-33.

103. Devesa SS, Donaldson J, Fears T. Graphical presentation of trends in rates. Am J Epidemiol. 1995;141(4):300-4.

104. Gladen BC, Rogan WJ. On graphing rate ratios. Am J Epidemiol. 1983;118(6):905-8.

105. Hebert JR, Miller DR. Plotting and discussion of rate ratios and relative risk estimates. J Clin Epidemiol. 1989;42(3):289-90.

106. Morgenstern H, Greenland S. Graphing ratio measures of effect. J Clin Epidemiol. 1990;43(6):539-42.

107. Robertson C, Boyle P. Age-period-cohort models of chronic disease rates. II: Graphical approaches. Stat Med. 1998;17(12):1325-39.

108. Tufte E. The visual display of quantitative information. Cheshire, CT: Graphics Press. 1983.

109. Doll R, Hill AB. A study of the aetiology of carcinoma of the lung. Br Med J. 1952;2(4797):1271-86.

110. Bray F, Ferlay J. Age standardization. IARC Sci Publ. 2014(164 Pt 1):112-5.

111. Armitage P, Doll R. The age distribution of cancer and a multi-stage theory of carcinogenesis. Br J Cancer. 1954;8(1):1-12.

112. Ahmad OE B-PC, Lopez AD, Murray CJL, Lozano R, Inoue M. Age standardization of rates: a new WHO standard GPE Discussion Paper Series: No. 31. Geneva: World Health Organization. . 2002.

113. Bray F, Guilloux A, Sankila R, Parkin DM. Practical implications of imposing a new world standard population. Cancer Causes Control. 2002;13(2):175-82.

114. Naing NN. Easy way to learn standardization : direct and indirect methods. Malays J Med Sci. 2000;7(1):10-5.

115. Doll R, Cook P. Summarizing indices for comparison of cancer incidence data. Int J Cancer. 1967;2(3):269-79.

116. Torre LA, Siegel RL, Ward EM, Jemal A. Global Cancer Incidence and Mortality Rates and Trends-An Update. Cancer Epidemiol Biomarkers Prev. 2016;25(1):16-27.

117. Chu KC, Baker SG, Tarone RE. A method for identifying abrupt changes in U.S. cancer mortality trends. Cancer. 1999;86(1):157-69.

118. Kim HJ, Fay MP, Feuer EJ, Midthune DN. Permutation tests for joinpoint regression with applications to cancer rates. Stat Med. 2000;19(3):335-51.

119. Kim HJ, Luo J, Kim J, Chen HS, Feuer EJ. Clustering of trend data using joinpoint regression models. Stat Med. 2014;33(23):4087-103.

120. Bray F, de Vries E. Non-identifiability and the age period cohort model: firm comprehension is an a Priori prerequisite. Ann Epidemiol. 2004;14(4):304-5; author reply 6-8.

121. Bray F, Richiardi L, Ekbom A, Forman D, Pukkala E, Cuninkova M, et al. Do testicular seminoma and nonseminoma share the same etiology? Evidence from an age-period-cohort analysis of incidence trends in eight European countries. Cancer Epidemiol Biomarkers Prev. 2006;15(4):652-8.

122. Bray I, Brennan P, Boffetta P. Recent trends and future projections of lymphoid neoplasms--a Bayesian age-period-cohort analysis. Cancer Causes Control. 2001;12(9):813-20.

123. Last JM. A dictionary of epidemiology (4th Ed). Oxford University Press. 2001.

124. Keyes KM, Utz RL, Robinson W, Li G. What is a cohort effect? Comparison of three statistical methods for modeling cohort effects in obesity prevalence in the United States, 1971-2006. Soc Sci Med. 2010;70(7):1100-8.

125. Kuh D, Smith GD. When is mortality risk determined? Historical insights into a current debate. Soc Hist Med. 1993;6(1):101-23.

126. Kermack WO, McKendrick AG, McKinlay PL. Death-rates in Great Britain and Sweden. Some general regularities and their significance. Int J Epidemiol. 2001;30(4):678-83.

127. McNally RJ, Alexander FE, Staines A, Cartwright RA. A comparison of three methods of analysis for age-period-cohort models with application to incidence data on non-Hodgkin's lymphoma. Int J Epidemiol. 1997;26(1):32-46.

128. Choi BC. Perspectives on epidemiologic surveillance in the 21st century. Chronic Dis Can. 1998;19(4):145-51.

129. Choi BC. The past, present, and future of public health surveillance. Scientifica (Cairo). 2012;2012:875253.

130. Choi BC, Pak AW. Lessons for surveillance in the 21st century: a historical perspective from the past five millennia. Soz Praventivmed. 2001;46(6):361-8.

131. Declich S, Carter AO. Public health surveillance: historical origins, methods and evaluation. Bull World Health Organ. 1994;72(2):285-304.

132. Eylenbosch WJ, Noah, N.D. "Historical aspects", in Surveillance in Health and Disease: Oxford University Press, Oxford, UK.; 1988.

133. Farr W. Vital statistics: memorial volume of selections from the reports and writings. 1885. Bull World Health Organ2000. p. 88-95.

134. Langmuir AD. William Farr: founder of modern concepts of surveillance. Int J Epidemiol. 1976;5(1):13-8.

135. Chadwick JM, W. On Airs, Waters, and Places. . Hippocratic Writings: Harmondsworth; 1978. p. 148-69.

136. Moro M, A, McCormick. . Surveillance for communicable disease. In: Eylenbosch WJ, Noah, N.D., editor. Surveillance in health and disease. Oxford, UK: Oxford University Press; 1988. p. 166-82.

137. Jones HW. John Graunt and His Bills of Mortality. Bull Med Libr Assoc. 1945;33(1):3-4.

138. Brownlee J. Historical Note on Farr's Theory of the Epidemic. Br Med J. 1915;2(2850):250-2.

139. Stocks P. Cancer registration and studies of incidence by surveys. Bull World Health Organ. 1959;20:697-715.

140. SYMPOSIUM on geographical pathology and demography of cancer, session held at Regent's Park College, Oxford, England July 29 to August, 5, 1950. Acta Unio Int Contra Cancrum. 1951;7(1 Spec. No.):1-152.

141. Davis W. IARC: 20 years old World Health. 1986:28-9.

142. Clemmesen J. Statistical Studies in the Aetiology of Malignant Neoplasms. I. Review and Results. Supplement 174. I. Acta Pathol Microbiol Scand Suppl. 1965;174:1-543.

143. Clemmesen J. Formal discussion of: the statistical approach to the etiology of malignant neoplasms. Cancer Res. 1965;25(8):1380-3.

144. Doll R, Muir C, Waterhouse J. Cancer incidence in five continents, vol. II. Berlin: Springer-Verlag (for UICC). 1970.

145. Parkin DM, Ferlay J, Curado MP, Bray F, Edwards B, Shin HR, et al. Fifty years of cancer incidence: CI5 I-IX. Int J Cancer. 2010;127(12):2918-27.

146. Jensen OM, Whelan S. Planning a cancer registry. IARC Sci Publ. 1991(95):22-8.

147. Dos Santos Silva I. The role of cancer registries. Cancer Epidemiology: Principles and Methods. Lyon, France: International Agency for Research on Cancer, WHO.; 1999. p. 442.

148. Young JL. The hospital-based cancer registry. IARC Sci Publ. 1991(95):177-84.

149. Gail M, Benichou J. Encyclopedia of epidemiologic methods. Wiley, Chichester. 2000:125-9.

150. Jedy-Agba EE, Curado MP, Oga E, Samaila MO, Ezeome ER, Obiorah C, et al. The role of hospital-based cancer registries in low and middle income countries-The Nigerian Case Study. Cancer Epidemiol. 2012;36(5):430-5.

151. Bray F, Znaor A, Cueva P, Korir A, Sawaminathan R, Ullrich A, et al. Planning and developing population-based cancer registration in low- and middle-income settings. Lyon Cedex 08, France: International Agency for Research on Cancer; 2014. Contract No.: IARC Technical Report No. 43.

152. Powell J. Cancer registration: principles and methods. Data sources and reporting. IARC Sci Publ. 1991(95):29-42.

153. Bullard J, Coleman MP, Robinson D, Lutz JM, Bell J, Peto J. Completeness of cancer registration: a new method for routine use. Br J Cancer. 2000;82(5):1111-6.

154. Das B, Clegg LX, Feuer EJ, Pickle LW. A new method to evaluate the completeness of case ascertainment by a cancer registry. Cancer Causes Control. 2008;19(5):515-25.

155. Inoue M, Tajima K, Inuzuka K, Tominaga S. The estimation of cancer incidence in Aichi Prefecture, Japan: use of degree of completeness of registration. J Epidemiol. 1998;8(1):60-4.

156. Parkin DM, Wabinga H, Nambooze S. Completeness in an African cancer registry. Cancer Causes Control. 2001;12(2):147-52.

157. Suwanrungruang K, Sriplung H, Attasara P, Temiyasathit S, Buasom R, Waisri N, et al. Quality of case ascertainment in cancer registries: a proposal for a virtual three-source capture-recapture technique. Asian Pac J Cancer Prev. 2011;12(1):173-8.

158. Tyczynski JE, Démaret E, Parkin DM. Standards and guidelines for cancer registration in Europe : the ENCR recommendations. (IARC technical publication ; no. 40) International Agency for Research on Cancer; 2003.

159. Goldberg J, Gelfand HM, Levy PS. Registry evaluation methods: a review and case study. Epidemiol Rev. 1980;2:210-20.

160. Donaldson L. Registering a need. BMJ. 1992;305(6854):597-8.

161. Zanetti R, Schmidtmann I, Sacchetto L, Binder-Foucard F, Bordoni A, Coza D, et al. Completeness and timeliness: Cancer registries could/should improve their performance. Eur J Cancer. 2015;51(9):1091-8.

162. Brewster D. Improving the quality of cancer registration data. J R Soc Med. 1995;88(5):268-71.

163. Crocetti E, Miccinesi G, Paci E, Zappa M. An application of the two-source capture-recapture method to estimate the completeness of the Tuscany Cancer Registry, Italy. Eur J Cancer Prev. 2001;10(5):417-23.

164. Skeet RG. Cancer registration: principles and methods. Quality and quality control. IARC Sci Publ. 1991(95):101-7.

165. Schmidtmann I. Estimating completeness in cancer registries--comparing capture-recapture methods in a simulation study. Biom J. 2008;50(6):1077-92.

166. Schmidtmann I, Blettner M. How do cancer registries in europe estimate completeness of registration? Methods Inf Med. 2009;48(3):267-71.

167. Robles SC, Marrett LD, Clarke EA, Risch HA. An application of capture-recapture methods to the estimation of completeness of cancer registration. J Clin Epidemiol. 1988;41(5):495-501.

168. Colonna M, Grosclaude P, Faivre J, Revzani A, Arveux P, Chaplain G, et al. Cancer registry data based estimation of regional cancer incidence: application to breast and colorectal cancer in French administrative regions. J Epidemiol Community Health. 1999;53(9):558-64.

169. Castro C, Bento MJ, Lunet N, Campos P. Assessing the completeness of cancer registration using suboptimal death certificate information. Eur J Cancer Prev. 2012;21(5):478-9.

170. Ajiki W, Tsukuma H, Oshima A. [Index for evaluating completeness of registration in population-based cancer registries and estimation of registration rate at the Osaka Cancer Registry between 1966 and 1992 using this index]. Nihon Koshu Eisei Zasshi. 1998;45(10):1011-7.

171. Kamo K, Kaneko S, Satoh K, Yanagihara H, Mizuno S, Sobue T. A mathematical estimation of true cancer incidence using data from population-based cancer registries. Jpn J Clin Oncol. 2007;37(2):150-5.

172. Hackl M, Waldhoer T. Estimation of completeness of case ascertainment of Austrian cancer incidence data using the flow method. Eur J Public Health. 2013;23(5):889-93.

173. Castro C. Estimating Completeness of Cancer Registration - An application to gastric cancer in the North Region Cancer Registry of Portugal [Masters]: Universidade do Porto; 2010.

174. Brenner H, Stegmaier C, Ziegler H. Estimating completeness of cancer registration: an empirical evaluation of the two source capture-recapture approach in Germany. J Epidemiol Community Health. 1995;49(4):426-30.

175. Azevedo-Silva F, Reis Rde S, Santos Mde O, Luiz RR, Pombo-de-Oliveira MS. Evaluation of childhood acute leukemia incidence and underreporting in Brazil by capture-recapture methodology. Cancer Epidemiol. 2009;33(6):403-5.

176. Silcocks PB, Robinson D. Completeness of ascertainment by cancer registries: putting bounds on the number of missing cases. J Public Health (Oxf). 2004;26(2):161-7.

177. Larsen IK, Smastuen M, Johannesen TB, Langmark F, Parkin DM, Bray F, et al. Data quality at the Cancer Registry of Norway: an overview of comparability, completeness, validity and timeliness. Eur J Cancer. 2009;45(7):1218-31.

178. Brenner H, Stegmaier C, Ziegler H. Estimating completeness of cancer registration in Saarland/Germany with capture-recapture methods. Eur J Cancer. 1994;30A(11):1659-63.

179. Tilling K. Capture-recapture methods--useful or misleading? Int J Epidemiol. 2001;30(1):12-4.

180. Tingulstad S, Halvorsen T, Norstein J, Hagen B, Skjeldestad FE. Completeness and accuracy of registration of ovarian cancer in the cancer registry of Norway. Int J Cancer. 2002;98(6):907-11.

181. Villard-Mackintosh L, Coleman MP, Vessey MP. The completeness of cancer registration in England: an assessment from the Oxford-FPA contraceptive study. Br J Cancer. 1988;58(4):507-11.

182. Pearce MS, Parker L. Childhood cancer registrations in the developing world: still more boys than girls. Int J Cancer. 2001;91(3):402-6.

183. Rutherford MJ, Moller H, Lambert PC. A comprehensive assessment of the impact of errors in the cancer registration process on 1- and 5-year relative survival estimates. Br J Cancer. 2013;108(3):691-8.

184. Burge F, Kockelbergh R. Closing the Gender Gap: Can We Improve Bladder Cancer Survival in Women? - A Systematic Review of Diagnosis, Treatment and Outcomes. Urol Int. 2016;97(4):373-9.

185. Dominguez-Gordillo A, Esparza-Gomez G, Garcia-Jimenez B, Cerero-Lapiedra R, Casado-Gomez I, Romero-Lastra P, et al. The pattern of lip cancer occurrence over the 1990-2011 period in public hospitals in Madrid, Spain. J Oral Pathol Med. 2016;45(3):202-10.

186. Fenner A. Bladder cancer: Bridging the gender gap in bladder cancer diagnoses. Nat Rev Urol. 2013;10(3):127.

187. Henning A, Wehrberger M, Madersbacher S, Pycha A, Martini T, Comploj E, et al. Do differences in clinical symptoms and referral patterns contribute to the gender gap in bladder cancer? BJU Int. 2013;112(1):68-73.

188. Juneja A, Adhikari T, Pandey A, Sharma S, Sehgal A. Share of tobacco related cancers: gender and time gaps-Indian scenario. J Clin Diagn Res. 2015;9(1):LC01-3.

189. Pang HH, Wang X, Stinchcombe TE, Wong ML, Cheng P, Ganti AK, et al. Enrollment Trends and Disparity Among Patients With Lung Cancer in National Clinical Trials, 1990 to 2012. J Clin Oncol. 2016.

190. Tabatabai MA, Kengwoung-Keumo JJ, Oates GR, Guemmegne JT, Akinlawon A, Ekadi G, et al. Racial and Gender Disparities in Incidence of Lung and Bronchus Cancer in the United States: A Longitudinal Analysis. PLoS One. 2016;11(9):e0162949.

191. Lapham R, Waugh NR. An audit of the quality of cancer registration data. Br J Cancer. 1992;66(3):552-4.

192. Lang K, Magi M, Aareleid T. Study of completeness of registration at the Estonian cancer registry. Eur J Cancer Prev. 2003;12(2):153-6.

193. Dickinson HO, Salotti JA, Birch PJ, Reid MM, Malcolm A, Parker L. How complete and accurate are cancer registrations notified by the National Health Service Central Register for England and Wales? J Epidemiol Community Health. 2001;55(6):414-22.

194. Barlow L, Westergren K, Holmberg L, Talback M. The completeness of the Swedish Cancer Register: a sample survey for year 1998. Acta Oncol. 2009;48(1):27-33.

195. Templeton AC, Bianchi A. Bias in an African cancer registry. Int J Cancer. 1972;10(1):186-93.

196. Templeton AC, Buxton E, Bianchi A. Cancer in Kyadondo County, Uganda, 1968-70. J Natl Cancer Inst. 1972;48(4):865-74.

197. Binagwaho A, Farmer PE, Nsanzimana S, Karema C, Gasana M, de Dieu Ngirabega J, et al. Rwanda 20 years on: investing in life. Lancet. 2014;384(9940):371-5.

198. Swaminathan R, Sankaranarayanan R. Under-diagnosis and under-ascertainment of cases may be the reasons for low childhood cancer incidence in rural India. Cancer Epidemiol. 2010;34(1):107-8.

199. Zakaria D. An examination of the NAACCR method of assessing completeness of case ascertainment using the Canadian Cancer Registry. Health Rep. 2013;24(8):3-13.

200. Shimakawa Y, Bah E, Wild CP, Hall AJ. Evaluation of data quality at the Gambia national cancer registry. Int J Cancer. 2013;132(3):658-65.

201. Bah E, Parkin DM, Hall AJ, Jack AD, Whittle H. Cancer in the Gambia: 1988-97. Br J Cancer. 2001;84(9):1207-14.

202. Monroe AC, Ricketts TC, Savitz LA. Cancer in rural versus urban populations: a review. J Rural Health. 1992;8(3):212-20.

203. Swaminathan R, Selvakumaran R, Esmy PO, Sampath P, Ferlay J, Jissa V, et al. Cancer pattern and survival in a rural district in South India. Cancer Epidemiol. 2009;33(5):325-31.

204. Howe H. Urban-rural Gradients in Cancer Incidence and Mortality in the United States Springfield, Illinois North American Association of Central Cancer Registries, Inc.; 2005.

205. Dimitrova N, Parkin DM. Data quality at the Bulgarian National Cancer Registry: An overview of comparability, completeness, validity and timeliness. Cancer Epidemiol. 2015;39(3):405-13.

206. Navarro C, Martos C, Ardanaz E, Galceran J, Izarzugaza I, Peris-Bonet R, et al. Population-based cancer registries in Spain and their role in cancer control. Ann Oncol. 2010;21 Suppl 3:iii3-13.

207. Tumino R, Ferretti S. Quality and completeness indices. Epidemiol Prev. 2004;28(2 Suppl):17-21.

208. Mathew A, Daniel CR, Ferrucci LM, Seth T, Devesa SS, George PS, et al. Assessment of follow-up, and the completeness and accuracy of cancer case ascertainment in three areas of India. Cancer Epidemiol. 2011;35(4):334-41.

209. Sakellariou D, Rotarou ES. Utilisation of cancer screening services by disabled women in Chile. PLoS One. 2017;12(5):e0176270.

210. Mandelblatt JS, Yabroff KR, Kerner JF. Equitable access to cancer services: A review of barriers to quality care. Cancer. 1999;86(11):2378-90.

211. Peters K, Cotton A. Barriers to breast cancer screening in Australia: experiences of women with physical disabilities. J Clin Nurs. 2015;24(3-4):563-72.

212. Ramjan L, Cotton A, Algoso M, Peters K. Barriers to breast and cervical cancer screening for women with physical disability: A review. Women Health. 2016;56(2):141-56.

213. McCusker J, Morrow GR. Factors related to the use of cancer early detection techniques. Prev Med. 1980;9(3):388-97.

214. Womeodu RJ, Bailey JE. Barriers to cancer screening. Med Clin North Am. 1996;80(1):115-33.

215. Parker L. Children's cancer in the developing world: where are the girls? Pediatr Hematol Oncol. 1998;15(2):99-103.

216. Parkin DM. The role of cancer registries in cancer control. Int J Clin Oncol. 2008;13(2):102-11.

217. Barnett MH, Williams DB, Day S, Macaskill P, McLeod JG. Progressive increase in incidence and prevalence of multiple sclerosis in Newcastle, Australia: a 35-year study. J Neurol Sci. 2003;213(1-2):1-6.

218. Ascherio A, Munger KL. Environmental risk factors for multiple sclerosis. Part II: Noninfectious factors. Ann Neurol. 2007;61(6):504-13.

219. Ascherio A, Munger KL. Environmental risk factors for multiple sclerosis. Part I: the role of infection. Ann Neurol. 2007;61(4):288-99.

220. Ramagopalan SV, Byrnes JK, Orton SM, Dyment DA, Guimond C, Yee IM, et al. Sex ratio of multiple sclerosis and clinical phenotype. Eur J Neurol. 2010;17(4):634-7.

221. Hernan MA, Hohol MJ, Olek MJ, Spiegelman D, Ascherio A. Oral contraceptives and the incidence of multiple sclerosis. Neurology. 2000;55(6):848-54.

222. Celius EG, Smestad C. Change in sex ratio, disease course and age at diagnosis in Oslo MS patients through seven decades. Acta Neurol Scand Suppl. 2009(189):27-9.

223. Koch-Henriksen N, Sorensen PS. The changing demographic pattern of multiple sclerosis epidemiology. Lancet Neurol. 2010;9(5):520-32.

224. Taylor BV, Pearson JF, Clarke G, Mason DF, Abernethy DA, Willoughby E, et al. MS prevalence in New Zealand, an ethnically and latitudinally diverse country. Mult Scler. 2010;16(12):1422-31.

225. Orton SM, Wald L, Confavreux C, Vukusic S, Krohn JP, Ramagopalan SV, et al. Association of UV radiation with multiple sclerosis prevalence and sex ratio in France. Neurology. 2011;76(5):425-31.

226. Morozova N, O'Reilly EJ, Ascherio A. Variations in gender ratios support the connection between smoking and Parkinson's disease. Mov Disord. 2008;23(10):1414-9.

227. Burn DJ. Sex and Parkinson's disease: a world of difference? J Neurol Neurosurg Psychiatry. 2007;78(8):787.

228. Moisan F, Kab S, Mohamed F, Canonico M, Le Guern M, Quintin C, et al. Parkinson disease male-to-female ratios increase with age: French nationwide study and meta-analysis. J Neurol Neurosurg Psychiatry. 2016;87(9):952-7.

229. Taylor KS, Cook JA, Counsell CE. Heterogeneity in male to female risk for Parkinson's disease. J Neurol Neurosurg Psychiatry. 2007;78(8):905-6.

230. van Vollenhoven RF. Sex differences in rheumatoid arthritis: more than meets the eye. BMC Med. 2009;7:12.

231. Lockshin MD. Invited review: sex ratio and rheumatic disease. J Appl Physiol (1985). 2001;91(5):2366-73.

232. Lockshin MD. Sex ratio and rheumatic disease. Isr Med Assoc J. 2001;3(7):511-6.

233. Lockshin MD. Sex ratio and rheumatic disease. Autoimmun Rev. 2002;1(3):162-7.

234. Lockshin MD. Sex ratio and rheumatic disease: excerpts from an Institute of Medicine report. Lupus. 2002;11(10):662-6.

235. Kvien TK, Uhlig T, Odegard S, Heiberg MS. Epidemiological aspects of rheumatoid arthritis: the sex ratio. Ann N Y Acad Sci. 2006;1069:212-22.

236. Palm O, Purinszky E. Women with early rheumatoid arthritis are referred later than men. Ann Rheum Dis. 2005;64(8):1227-8.

237. Markle JG, Fish EN. SeXX matters in immunity. Trends Immunol. 2014;35(3):97-104.

238. Roved J, Westerdahl H, Hasselquist D. Sex differences in immune responses: Hormonal effects, antagonistic selection, and evolutionary consequences. Horm Behav. 2017;88:95-105.

239. vom Steeg LG, Klein SL. SeXX Matters in Infectious Disease Pathogenesis. PLoS Pathog. 2016;12(2):e1005374.

240. Zuk M, McKean KA. Sex differences in parasite infections: patterns and processes. Int J Parasitol. 1996;26(10):1009-23.

241. Guerra-Silveira F, Abad-Franch F. Sex bias in infectious disease epidemiology: patterns and processes. PLoS One. 2013;8(4):e62390.

242. Green MS. The male predominance in the incidence of infectious diseases in children: a postulated explanation for disparities in the literature. Int J Epidemiol. 1992;21(2):381-6.

243. Sasaki S, Zhang XH, Kesteloot H. Dietary sodium, potassium, saturated fat, alcohol, and stroke mortality. Stroke. 1995;26(5):783-9.

244. Zhang J, Sasaki S, Amano K, Kesteloot H. Fish consumption and mortality from all causes, ischemic heart disease, and stroke: an ecological study. Prev Med. 1999;28(5):520-9.

245. Zhang XH, Sasaki S, Kesteloot H. Changes in the sex ratio of stroke mortality in the period of 1955 through 1990. Stroke. 1995;26(10):1774-80.

246. Lee KT, Thomas WA. Factors associated with changing sex ratio of myocardial infarction; study with special reference to the disproportionate rise in incidence of the disease among older women. AMA Arch Intern Med. 1956;98(1):80-3.

247. Rosengren A, Thelle DS, Koster M, Rosen M. Changing sex ratio in acute coronary heart disease: data from Swedish national registers 1984-99. J Intern Med. 2003;253(3):301-10.

248. Zhang J, Temme EH, Kesteloot H. Fish consumption is inversely associated with male lung cancer mortality in countries with high levels of cigarette smoking or animal fat consumption. Int J Epidemiol. 2000;29(4):615-21.

249. Nicholson WJ, Davis DL. Analyses of changes in the ratios of male-to-female cancer mortality. A hypothesis-generating exercise. Ann N Y Acad Sci. 1990;609:290-7; discussion 7-9.

250. Radkiewicz C, Johansson ALV, Dickman PW, Lambe M, Edgren G. Sex differences in cancer risk and survival: A Swedish cohort study. Eur J Cancer. 2017;84:130-40.

251. Ames BN. Mutagenesis and carcinogenesis: endogenous and exogenous factors. Environ Mol Mutagen. 1989;14 Suppl 16:66-77.

252. Greenwald HP. Who Survives Cancer? Berkeley: University of California Press. http://ark.cdlib.org/ark:/13030/ft9b69p365/

1992.

253. Lutz WK, Fekete T. Endogenous and exogenous factors in carcinogenesis: limits to cancer prevention. Int Arch Occup Environ Health. 1996;68(2):120-5.

254. Rosenberg MS, Sokal RR, Oden NL, DiGiovanni D. Spatial autocorrelation of cancer in Western Europe. Eur J Epidemiol. 1999;15(1):15-22.

255. Oksuzyan A, Juel K, Vaupel JW, Christensen K. Men: good health and high mortality. Sex differences in health and aging. Aging Clin Exp Res. 2008;20(2):91-102.

256. Wingard DL. The sex differential in morbidity, mortality, and lifestyle. Annu Rev Public Health. 1984;5:433-58.

257. Naugler WE, Sakurai T, Kim S, Maeda S, Kim K, Elsharkawy AM, et al. Gender disparity in liver cancer due to sex differences in MyD88-dependent IL-6 production. Science. 2007;317(5834):121-4.

258. Graham H. Smoking prevalence among women in the European community 1950-1990. Soc Sci Med. 1996;43(2):243-54.

259. Salameh P, Khayat G, Waked M. Lower prevalence of cigarette and waterpipe smoking, but a higher risk of waterpipe dependence in Lebanese adult women than in men. Women Health. 2012;52(2):135-50.

260. Travasso C. Smoking prevalence among Indian women doubled in five years, study finds. BMJ. 2014;349:g7512.

261. Wennergren G, Ekerljung L, Alm B, Bjerg A, Lotvall J, Lundback B. Alarmingly high prevalence of smoking and symptoms of bronchitis in young women in Sweden: a population-based questionnaire study. Prim Care Respir J. 2013;22(2):214-20.

262. Nicholson WJ. IARC evaluations in the light of limitations of human epidemiologic data. Ann N Y Acad Sci. 1988;534:44-54.

263. Field A. Discovering statistics using IBM SPSS statistics. Fourth Edition. Carmichael M, editor: SAGE Publications. London; 2013.

264. Twisk JWR. Applied multilevel analysis: a practical guide. First ed: Cambridge University Press. Cambridge; 2006.

265. WHO report on the global tobacco epidemic 2008. World Health Organization, Geneva, Switzerland. Accessed on March 31, 2017 [Available from: http://www.who.int/tobacco/mpower/2008/en/].

266. Al-Ahmadi K, Al-Zahrani A. Spatial autocorrelation of cancer incidence in Saudi Arabia. Int J Environ Res Public Health. 2013;10(12):7207-28.

267. Chen Y. Spatial Autocorrelation Approaches to Testing Residuals from Least Squares Regression. PLoS One. 2016;11(1):e0146865.

268. Moran PA. Notes on continuous stochastic phenomena. Biometrika. 1950;37(1-2):17-23.

269. Chen Y. New approaches for calculating Moran's index of spatial autocorrelation. PLoS One. 2013;8(7):e68336.

270. Free World Cities Database . Open data license for MaxMind World Cities and Postal Code Databases. Accessed on November 1, 2015 [https://www.maxmind.com/en/free-world-cities-database ; .

271. Chen YG. On the four types of weight functions for spatial contiguity matrix. . Letters in Spatial Resource and Sciences. 2012;5(2):65-72.

272. Giovino GA. Epidemiology of tobacco use in the United States. Oncogene. 2002;21(48):7326-40.

273. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA Cancer J Clin. 2016;66(1):7-30.

274. Marley AR, Nan H. Epidemiology of colorectal cancer. Int J Mol Epidemiol Genet. 2016;7(3):105-14.

275. Harris JE. Cigarette smoking among successive birth cohorts of men and women in the United States during 1900-80. J Natl Cancer Inst. 1983;71(3):473-9.

276. Schmid D, Ricci C, Behrens G, Leitzmann MF. Adiposity and risk of thyroid cancer: a systematic review and meta-analysis. Obes Rev. 2015;16(12):1042-54.

277. O'Grady TJ, Gates MA, Boscoe FP. Thyroid cancer incidence attributable to overdiagnosis in the United States 1981-2011. Int J Cancer. 2015;137(11):2664-73.

278. Fitzmaurice C, Dicker D, Pain A, Hamavid H, Moradi-Lakeh M, MacIntyre MF, et al. The Global Burden of Cancer 2013. JAMA Oncol. 2015;1(4):505-27.

279. Wild C, Brennan P, Plummer M, Bray F, Straif K, Zavadil J. Cancer risk: role of chance overstated. Science. 2015;347(6223):728.

280. Epstein SS, Swartz JB. Fallacies of lifestyle cancer theories. Nature. 1981;289(5794):127-30.

281. Tomasetti C, Vogelstein B. Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. Science. 2015;347(6217):78-81.

282. Rothchild I. Induction, deduction, and scientific method. Society for the Study of Reproduction, Inc. 2006:1-11.

283. Kirchoff H, Rigdon RH. Cancer of the lung-the sex ratio; a review of the problem. Tex Rep Biol Med. 1959;17(1):29-48.

284. Walshe WH. A practical treatise on diseases of lungs. London: James Watson; 1871.

285. Dorn HF, Cutler SJ. Morbidity from cancer in the United States. I. Variation in incidence by age, sex, race, marital status, and geographic region. Public Health Monogr. 1955;29:1-121.

286. Dorn HF. Methods of measuring incidence and prevalence of disease. Am J Public Health Nations Health. 1951;41(3):271-8.

287. Hanbury WJ. Bronchogenic Carcinoma in Women. Thorax. 1964;19:338-42.

288. Bross IJ, Cornfield J, O'Donnell WE, Wynder EL. Lung cancer in women; a study of environmental factors. N Engl J Med. 1956;255(24):1111-21.

289. Flamant R, Lasserre O, Lazar P, Leguerinais J, Denoix P, Schwartz D. Differences in Sex Ratio According to Cancer Site and Possible Relationship with Use of Tobacco and Alcohol. Review of 65,000 Cases. J Natl Cancer Inst. 1964;32:1309-16.

290. La Vecchia C, Levi F. Sex differentials in Swiss cancer mortality. Soz Praventivmed. 1988;33(3):140-3.

291. Levi F, La Vecchia C, Lucchini F, Negri E. Trends in cancer mortality sex ratios in Europe, 1950-1989. World Health Stat Q. 1992;45(1):117-64.

292. Hebert JR, Daguise VG, Hurley DM, Wilkerson RC, Mosley CM, Adams SA, et al. Mapping cancer mortality-to-incidence ratios to illustrate racial and sex disparities in a high-risk population. Cancer. 2009;115(11):2539-52.

293. Karim-Kos HE, de Vries E, Soerjomataram I, Lemmens V, Siesling S, Coebergh JW. Recent trends of cancer in Europe: a combined approach of incidence, survival and mortality for 17 cancer sites since the 1990s. Eur J Cancer. 2008;44(10):1345-89.

294. Micheli A, Ciampichini R, Oberaigner W, Ciccolallo L, de Vries E, Izarzugaza I, et al. The advantage of women in cancer survival: an analysis of EUROCARE-4 data. Eur J Cancer. 2009;45(6):1017-27.

295. Micheli A, Mariotto A, Giorgi Rossi A, Gatta G, Muti P. The prognostic role of gender in survival of adult cancer patients. EUROCARE Working Group. Eur J Cancer. 1998;34(14 Spec No):2271-8.

296. Edgren G, Adami HO, Weiderpass E, Nyren O. A global assessment of the oesophageal adenocarcinoma epidemic. Gut. 2013;62(10):1406-14.

297. McMichael AJ, Potter JD. Do intrinsic sex differences in lower alimentary tract physiology influence the sex-specific risks of bowel cancer and other biliary and intestinal diseases? Am J Epidemiol. 1983;118(5):620-7.

298. Stephenson WT, Barnes DE, Holmes FF, Norris CW. Gender influences subsite of origin of laryngeal carcinoma. Arch Otolaryngol Head Neck Surg. 1991;117(7):774-8.

299. Sipponen P, Correa P. Delayed rise in incidence of gastric cancer in females results in unique sex ratio (M/F) pattern: etiologic hypothesis. Gastric Cancer. 2002;5(4):213-9.

300. Mathieu LN, Kanarek NF, Tsai HL, Rudin CM, Brock MV. Age and sex differences in the incidence of esophageal adenocarcinoma: results from the Surveillance, Epidemiology, and End Results (SEER) Registry (1973-2008). Dis Esophagus. 2014;27(8):757-63.

301. Danaei G, Vander Hoorn S, Lopez AD, Murray CJ, Ezzati M, Comparative Risk Assessment collaborating g. Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors. Lancet. 2005;366(9499):1784-93.

302. Chaturvedi HK, Phukan RK, Zoramtharga K, Hazarika NC, Mahanta J. Tobacco use in Mizoram, India: sociodemographic differences in pattern. Southeast Asian J Trop Med Public Health. 1998;29(1):66-70.

303. Phukan RK, Zomawia E, Narain K, Hazarika NC, Mahanta J. Tobacco use and stomach cancer in Mizoram, India. Cancer Epidemiol Biomarkers Prev. 2005;14(8):1892-6.

304. Sinha DN, Gupta PC, Pednekar M. Tobacco water: a special form of tobacco use in the Mizoram and Manipur states of India. Natl Med J India. 2004;17(5):245-7.

305. Cigarette smoking among adults-United States, 2006. MMWR Morb Mortal Wkly Rep. 2007;56:1157-61.

306. Corley DA, Kubo A, Zhao W. Abdominal obesity and the risk of esophageal and gastric cardia carcinomas. Cancer Epidemiol Biomarkers Prev. 2008;17(2):352-8.

307. Ter RB. Gender differences in gastroesophageal reflux disease. J Gend Specif Med. 2000;3(2):42-4.

308. Banki F, Demeester SR, Mason RJ, Campos G, Hagen JA, Peters JH, et al. Barrett's esophagus in females: a comparative analysis of risk factors in females and males. Am J Gastroenterol. 2005;100(3):560-7.

309. Sugerman H, Windsor A, Bessos M, Wolfe L. Intra-abdominal pressure, sagittal abdominal diameter and obesity comorbidity. J Intern Med. 1997;241(1):71-9.

310. Wynder EL, Mabuchi K. Cancer of the gastrointestinal tract. Etiological and environmental factors. JAMA. 1973;226(13):1546-8.

311. De Stefani E, Correa P, Oreggia F, Leiva J, Rivero S, Fernandez G, et al. Risk factors for laryngeal cancer. Cancer. 1987;60(12):3087-91.

312. Markou K, Christoforidou A, Karasmanis I, Tsiropoulos G, Triaridis S, Constantinidis I, et al. Laryngeal cancer: epidemiological data from Nuorthern Greece and review of the literature. Hippokratia. 2013;17(4):313-8.

313. Wynder EL, Covey LS, Mabuchi K, Mushinski M. Environmental factors in cancer of the larynx: a second look. Cancer. 1976;38(4):1591-601.

314. Yang PC, Thomas DB, Daling JR, Davis S. Differences in the sex ratio of laryngeal cancer incidence rates by anatomic subsite. J Clin Epidemiol. 1989;42(8):755-8.

315. Berrino F, Richiardi L, Boffetta P, Esteve J, Belletti I, Raymond L, et al. Occupation and larynx and hypopharynx cancer: a job-exposure matrix approach in an international case-control study in France, Italy, Spain and Switzerland. Cancer Causes Control. 2003;14(3):213-23.

316. Boffetta P, Richiardi L, Berrino F, Esteve J, Pisani P, Crosignani P, et al. Occupation and larynx and hypopharynx cancer: an international case-control study in France, Italy, Spain, and Switzerland. Cancer Causes Control. 2003;14(3):203-12.

317. Cattaruzza MS, Maisonneuve P, Boyle P. Epidemiology of laryngeal cancer. Eur J Cancer B Oral Oncol. 1996;32B(5):293-305.

318. Muscat JE, Wynder EL. Tobacco, alcohol, asbestos, and occupational risk factors for laryngeal cancer. Cancer. 1992;69(9):2244-51.

319. Brewster DH, Bhatti LA. Increasing incidence of squamous cell carcinoma of the anus in Scotland, 1975-2002. Br J Cancer. 2006;95(1):87-90.

320. Johnson LG, Madeleine MM, Newcomer LM, Schwartz SM, Daling JR. Anal cancer incidence and survival: the surveillance, epidemiology, and end results experience, 1973-2000. Cancer. 2004;101(2):281-8.

321. McNally RJ, Blakey K, James PW, Gomez Pozo B, Basta NO, Hale J. Increasing incidence of thyroid cancer in Great Britain, 1976-2005: age-period-cohort analysis. Eur J Epidemiol. 2012;27(8):615-22.

322. Nagataki S, Nystrom E. Epidemiology and primary prevention of thyroid cancer. Thyroid. 2002;12(10):889-96.

323. Nasseri K. Thyroid cancer in the Middle Eastern population of California. Cancer Causes Control. 2008;19(10):1183-91.

324. Roti E, degli Uberti EC, Bondanelli M, Braverman LE. Thyroid papillary microcarcinoma: a descriptive and meta-analysis study. Eur J Endocrinol. 2008;159(6):659-73.

325. Schofield C. An annotated bibliography of source material for basal metabolic rate data. Hum Nutr Clin Nutr. 1985;39 Suppl 1:42-91.

326. Schofield WN. Predicting basal metabolic rate, new standards and review of previous work. Hum Nutr Clin Nutr. 1985;39 Suppl 1:5-41.

327. Albanes D, Winick M. Are cell number and cell proliferation risk factors for cancer? J Natl Cancer Inst. 1988;80(10):772-4.

328. Scelo G, Li P, Chanudet E, Muller DC. Variability of Sex Disparities in Cancer Incidence over 30 Years: The Striking Case of Kidney Cancer. Eur Urol Focus. 2017.

329. Wang TC, Szabo E. Implications of the "Bad Luck" Explanation of Cancer Risk for the Field of Cancer Prevention. Cancer Prev Res (Phila). 2015;8(9):761.

330. Rozhok AI, Wahl GM, DeGregori J. A Critical Examination of the "Bad Luck" Explanation of Cancer Risk. Cancer Prev Res (Phila). 2015;8(9):762-4.

331. Cao S, Zhang C, Xu Y. Somatic mutations may not be the primary drivers of cancer formation. Int J Cancer. 2015;137(11):2762-5.

332. Diez Roux AV. The study of group-level factors in epidemiology: rethinking variables, study designs, and analytical approaches. Epidemiol Rev. 2004;26:104-11.

333. Borchmann S, Muller H, Engert A. Hodgkin Lymphoma has a seasonal pattern of incidence and mortality that depends on latitude. Sci Rep. 2017;7(1):14903.

334. Freedman ND, Silverman DT, Hollenbeck AR, Schatzkin A, Abnet CC. Association between smoking and risk of bladder cancer among men and women. JAMA. 2011;306(7):737-45.

335. Cote ML, Colt JS, Schwartz KL, Wacholder S, Ruterbusch JJ, Davis F, et al. Cigarette smoking and renal cell carcinoma risk among black and white Americans: effect modification by hypertension and obesity. Cancer Epidemiol Biomarkers Prev. 2012;21(5):770-9.

336. Ng M, Freeman MK, Fleming TD, Robinson M, Dwyer-Lindgren L, Thomson B, et al. Smoking prevalence and cigarette consumption in 187 countries, 1980-2012. JAMA. 2014;311(2):183-92.

337. Non Communicable Disease Risk Factor Collaboration. Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19.2 million participants. Lancet. 2016;387(10026):1377-96.

338. Haas NB, Nathanson KL. Hereditary kidney cancer syndromes. Adv Chronic Kidney Dis. 2014;21(1):81-90.

339. Purdue MP, Johansson M, Zelenika D, Toro JR, Scelo G, Moore LE, et al. Genome-wide association study of renal cell carcinoma identifies two susceptibility loci on 2p21 and 11q13.3. Nat Genet. 2011;43(1):60-5.

340. Scelo G, Purdue MP, Brown KM, Johansson M, Wang Z, Eckel-Passow JE, et al. Genome-wide association study identifies multiple risk loci for renal cell carcinoma. Nat Commun. 2017;8:15724.

341. Wu X, Scelo G, Purdue MP, Rothman N, Johansson M, Ye Y, et al. A genome-wide association study identifies a novel susceptibility locus for renal cell carcinoma on 12p11.23. Hum Mol Genet. 2012;21(2):456-62.

342. Mohr SB, Gorham ED, Garland CF, Grant WB, Garland FC, Cuomo RE. Are low ultraviolet B and vitamin D associated with higher incidence of multiple myeloma? J Steroid Biochem Mol Biol. 2015;148:245-52.

343. Fearon ER. Human cancer syndromes: clues to the origin and nature of cancer. Science. 1997;278(5340):1043-50.

344. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med. 2000;343(2):78-85.

345. Grady D. Cancer's random assault. New York Times. January 6, 2015:Section D1.

346. Johnson G. Random Chance's role in cancer. New York Times. 2015, January 20:Section D:5.

347. Couzin-Frankel J. Bad luck and cancer: a science reporter's reflections on controversial story. Science http://wwwsciencemagorg/news/2015/01/bad-luck-and-cancer-science-reporter-s-reflections-controversial-story. 2015.

348. Couzin-Frankel J. Science Communication. Backlash greets 'bad luck' cancer study and coverage. Science. 2015;347(6219):224.

349. Ashford NA, Bauman P, Brown HS, Clapp RW, Finkel AM, Gee D, et al. Cancer risk: role of environment. Science. 2015;347(6223):727.

350. Gotay C, Dummer T, Spinelli J. Cancer risk: prevention is crucial. Science. 2015;347(6223):728.

351. O'Callaghan M. Cancer risk: accuracy of literature. Science. 2015;347(6223):729.

352. Potter JD, Prentice RL. Cancer risk: tumors excluded. Science. 2015;347(6223):727.

353. Song M, Giovannucci EL. Cancer risk: many factors contribute. Science. 2015;347(6223):728-9.

354. Regitz-Zagrosek V. Sex and gender differences in symptoms of myocardial ischaemia. Eur Heart J. 2011;32(24):3064-6.

355. Regitz-Zagrosek V. Sex and gender differences in health. Science & Society Series on Sex and Science. EMBO Rep. 2012;13(7):596-603.

356. Parkin DM, Stiller CA, Draper GJ, Bieber CA. The international incidence of childhood cancer. Int J Cancer. 1988;42(4):511-20.

357. Curado MP, Voti L, Sortino-Rachou AM. Cancer registration data and quality indicators in low and middle income countries: their interpretation and potential use for the improvement of cancer care. Cancer Causes Control. 2009;20(5):751-6.

358. Harper S, Lynch J, Meersman SC, Breen N, Davis WW, Reichman ME. An overview of methods for monitoring social disparities in cancer with an example using trends in lung cancer incidence by area-socioeconomic position and race-ethnicity, 1992-2004. Am J Epidemiol. 2008;167(8):889-99.

359. Bobak M. Relative and absolute gender gap in all-cause mortality in Europe and the contribution of smoking. Eur J Epidemiol. 2003;18(1):15-8.

360. Waldron I. Sex differences in human mortality: the role of genetic factors. Soc Sci Med. 1983;17(6):321-33.

Appendices

Appendix 1: Example of computations of CIR & ASIR for melanoma in Denmark in males, 1998-2002.

Stratum-Age group	*No. of cases*	*Person-yrs at risk*	*Age specific incidence per 100,000*	*Standard world population*	*Expected cases in standard population*
i	d_i	y_i	d_i / y_i	w_i	$d_i w_i / y_i$
0-4	0	869 668	0.00	12 000	0.00
5-9	0	878 535	0.00	10 000	0.00
10-14	0	779 915	0.00	9 000	0.00
15-19	7	723 313	0.97	9 000	0.09
20-24	23	840 592	2.74	8 000	0.22
25-29	59	961 907	6.13	8 000	0.49
30-34	83	1 042 749	7.96	6 000	0.48
35-39	117	1 041 800	11.23	6 000	0.67
40-44	153	953 731	16.04	6 000	0.96
45-49	188	930 210	20.21	6 000	1.21
50-54	262	990 187	26.46	5 000	1.32
55-59	303	871 969	34.75	4 000	1.39
60-64	272	649 423	41.88	4 000	1.68
65-69	223	523 486	42.60	3 000	1.28
70-74	217	439 676	49.35	2 000	0.99
75-79	162	342 435	47.31	1 000	0.47
80-84	119	210 084	56.64	500	0.28
85+	115	140 073	82.10	500	0.41
Total	**D=2303**	**Y=13 189 753**	**D/Y=17.46**	**100 000**	**11.94**

Appendix 2: Description of interpretation of Fixed-Effect Models *a* and *b*, and Random-Intercept Model *c*.

i : time period and $i = 1,\ldots,N$

j : registry and $j = 1,\ldots,n$

k: cancer type and $k = 1,\ldots,K$

Fixed Effect Model *a*

$$SR_i^k = \beta_0^k + \beta_1^k year_i + \varepsilon_i^k ,$$

$$where\ \varepsilon_i^k \sim N\ (0, (\delta^k)^2)$$

$Interpretation$:

β_0^k	This is the intercept and signifies what the average SR for the selected cancer in the baseline year (which can be 1953 or 1983 depending on the scenario A or B/C, respectively).
$\beta_1^k year_i$	This is the slope of linear time trend that applies to all registries and indicates the average annual rate of increase or decrease (depending on positive or negative sign) of SR of the selected cancer.

Fixed effect Model *b*

$$SR_i^k = \beta_0^k + \beta_1^k year_i + \varepsilon_i^k + \beta_2^k(registry_1) + \beta_3^k(registry_2), \ldots + \beta_n^k(registry_n)$$

$$where\ \varepsilon_i^k \sim N\ (0, (\delta^k)^2)$$

$Interpretation$:

β_0^k	This is the intercept of the *reference registry* (out of total of 113 or 76 depending on scenario A or B/C, respectively) in the model, and indicates the average value of the SR for the selected cancer in the baseline year.
$\beta_1^k year_i$	The interpretation is same as model *a*
$\beta_n^k(registry_n)$	In this model, depending on the scenario, there are 112 or 75 intercepts (β_n^k) each indicating how much on average the sex ratio is higher or lower than the reference registry (therefore β_n^k in each registry is the mean difference or the differences in the intercept).

Appendix 2 cont…..

Random Intercept model *c*

$SR_{ij}^{k} = \beta_0^k + \beta_1^k year_i + u_j^k + \varepsilon_{ij}^k$,

$where\ \varepsilon_{ij}^k \sim N\ (0, (\delta^k)^2)$, and

u_j^k is the random variations of intercepts of SR among registries and it has a variance τ_1^{k2}

$Interpretation$

β_0^k	This is the intercept and implies what SR for selected cancer on average is in baseline year.
$\beta_1^k year_i$	The average annual rate of increase or decrease in SR of selected cancer (while keeping slopes of populations constant) and this is provided with statistically significant testing (p value and 95% CI).
τ_1^{k2}	Random effect u_j^k has a single variance known as τ_1^{k2}. Therefore u_j^k indicates random variations in populations i.e., every population has a different SR at baseline year and has a single variance τ_1^{k2} also known as *inter-country variation*.
$(\delta^k)^2$	ε_{ij}^k has a variance$(\delta^k)^2$ δ^2. This is *intra-regional variation*.

Random intercept and Random slope model *d*

$$SR_{ij}^{k} = \beta_0^k + \beta_1^k year_i + u_{1j}^k + u_{2j}^k(year_i) + \varepsilon_{ij}^k , \quad where\ \varepsilon_{ij}^k \sim N\ (0, (\delta^k)^2)$$

$k =$ cancer type; $i =$ index of calender year; $j =$ registry

$\varepsilon_{ij}^k \sim N\ (0, (\delta^k)^2)$ *i*s the assumption that the errors are normally

and independantly distributed with mean 0 and variance $(\delta^k)^2$

and $u_{1j}^k \sim N\left(0, \tau_1^{k2}\right)$ for registry j ;

$u_{2j}^k \sim N\left(0, \tau_2^{k2}\right)$ for registry j ;

$$Cov\left(u_{1j}^k, u_{2j}^k\right) = \tau_{12}^k$$

u_{1j}^k is the random variations of intercepts of SR among registries and it has a variance τ_1^{k2};

and u_{2j}^k is the random variations of slopes and it has a variance τ_2^{k2}

The random intercepts and slopes are correlated $Cov\left(u_{1j}^k, u_{2j}^k\right) = \tau_{12}^k$

Appendix 2 cont…..

Interpretannation

β_0^k	This is the intercept and implies what SR for selected cancer on average is in baseline year.
$\beta_1^k year_i$	The average annual rate of increase or decrease in SR of selected cancer (while keeping slopes of populations constant).
τ_1^{k2}	Random effect u_j^k has a single variance known as τ_1^{k2}. Therefore u_j^k indicates random variations in populations i.e., every population has a different SR at baseline year and has a single variance τ_1^{k2} also known as *inter-country variation*.
τ_2^{k2}	Random effect u_{2j}^k has a single variance known as τ_2^{k2} .Therefore u_{2j}^k indicates random variations in changes in SR over years i.e., SR of every registry/region changes differently over years (also called variation of slopes).
$(\delta^k)^2$	ε_{ij}^k has a variance$(\delta^k)^2$. This is *intra-regional variation*.
τ_{12}^k	This is covariance between random intercepts and random slopes of every registry $Cov\left(u_{1j}^k, u_{2j}^k\right)$. A value of with a negative sign indicates that populations with high SR can have steeper decline in SR over years and vice versa.

Appendix 3: An example of calculations of distances from the latitude and longitude

1	A	B	C	D	E	F
2	**City**	**Latitude**	**Longitude**	**Distance ft**	**Distance m**	**Distance km**
3	China Beijing	39.92	116.38	1427134.49	434990.59	434.99
4	China Cixian	36.35	114.37	3689558.07	1124577.30	1124.57
5	China Haining	30.40	120.53	6481508.01	1975563.64	1975.56
6	China Hong K	22.28	114.15	612435.77	186670.42	186.67
7	China Jiashan	40.78	118.28	3593147.41	1095191.33	1095.19
8	China Jiaxing	30.75	120.75	4660938.37	1420654.01	1420.65
9	China Macao	31.90	104.10	6206331.84	1891689.94	1891.68
10	China Nangang	23.10	113.54	3347757.58	1020396.51	1020.39
11	China Qidong	31.81	121.65	3568466.55	1087668.60	1087.66
12	China Shanghai	31.04	121.39	3471665.32	1058163.59	1058.16

Computations of pairwise distances in excel (e.g., for distance between Beijing and Cixian

Distance in feet:

=(3963*ACOS(COS(RADIANS(90-B2))*COS(RADIANS(90-B3))+SIN(RADIANS(90-B2))*SIN(RADIANS(90-B3))*COS(RADIANS(C2-C3))))*5280

Conversion of Distance from feet into meters:

=CONVERT(D2,"ft","m")

Conversion of distance meters into kilometers:

=E2/1000

Appendix 4: Cancer types with their overall sex ratio magnitude (SRm) and variance (SRv) along with populations with highest and lowest SRm for incidence period of (a) 1973-1977, 1988-92, and 2003-07

(a) 1973-1977 – Cancer Incidence in Five Continents Volume IV

Cancer*	Overall SR		Populations with highest & lowest SRm			
	SRm	(SRv)	Highest SRm		Lowest SRm	
Lip	10.66	(90.01)	Spain Zaragoza	59.00	Senegal Dakar	0.33
			USA Iowa	46.00	India Mumbai	0.50
			Australia South	25.25	Colombia Cali	0.75
Tongue	3.55	(13.74)	Slovenia	30.00	China Shanghai	0.75
			Hungary Szabolcs-Szatmar	14.00	Finland	1.17
			Italy Varese	12.67	Scotland North	1.33
Salivary glands	1.34	(0.37)	Spain Navarra	3.00	Canada NF	0.44
			Canada Manitoba	3.00	Poland Nowy Sacz	0.50
			Australia South	2.50	UK Scotland North East	0.67
Mouth	3.29	(9.69)	Slovenia	15.50	Senegal Dakar	0.77
			Italy Varese	13.67	India Mumbai	1.00
			France Bas-Rhin	12.00	Colombia Cali	1.11
Pharynx	4.26	(20.88)	France Doubs	31.00	Poland Nowy Sacz	0.89
			Italy Varese	26.33	Israel	1.50
			France Doubs	20.27	UK England Mersey	1.55
Larynx	13.96	(344.94)	France Bas-Rhin	112.00	UK Scotland East	3.50
			Hungary County Vas	13.78	UK England Mersey	4.00
			Spain Navarra	12.93	UK Scotland West	4.33
Lung	5.25	(9.30)	France Doubs	15.11	Canada NWT & Yukon	2.08
			Germany Saarland	13.78	China Hong Kong	2.37
			France Bas-Rhin	12.93	Cuba	2.63

(a) 1973-77 cont…	SRm	(SRv)	Highest SRm		Lowest SRm	
Oesophagus	4.08	(12.31)	France Bas-Rhin	21.25	Senegal Dakar	1.00
			Hungary County Vas	20.00	India Poona	1.26
			France Doubs	18.57	Finland	1.29
Nose & sinuses	2.03	(1.62)	France Doubs	6.75	Hungary Szabolcs-Szatmar	0.60
			UK Scotland North	5.00	Switzerland Neuchatel	0.71
			Spain Zaragoza	5.00	Canada Maritime	0.75
Stomach	2.19	(0.21)	Switzerland Neuchatel	4.53	India Poona	1.14
			Canada NWT & Yukon	4.45	India Mumbai	1.64
			France Doubs	3.21	Israel	1.67
Small intestine	1.46	(0.52)	Italy Varese	4.00	Colombia Cali	0.50
			France Bas-Rhin	3.33	Poland Katowice	0.50
			Poland Cracow	3.00	Canada Maritime	0.67
Colon	1.14	(0.05)	Romania Cluj	1.57	Canada NWT & Yukon	0.65
			Switzerland Vaud	1.43	Cuba	0.77
			Poland Cracow	1.41	Colombia Cali	0.83
Rectum & anus	1.56	(0.10)	Canada NWT & Yukon	3.01	Cuba	1.05
			France Bas-Rhin	2.47	Switzerland Neuchatel	1.06
			Poland Cieszyn	2.02	Israel	1.10
Liver	2.54	(1.55)	Canada NWT & Yukon	7.46	Canada NWT & Yukon	0.38
			France Bas-Rhin	7.00	Spain Navarra	0.83
			Netherlands Antilles	4.83	Poland Warsaw Rural	0.91
Gallbladder	0.78	(0.18)	UK Scotland North	1.43	Hugary Szabolcs-Szatmar	0.21
			China Hong Kong	1.11	France Doubs	0.29
			UK England Oxford	1.00	Romania Cluj	0.31
Pancreas	1.63	(0.25)	Netherlands, Antilles	3.58	Canada NWT & Yukon	0.32
			France Bas-Rhin	2.86	Jamaica Kingston & St Andrew	0.72
			Poland Cieszyn	2.78	Senegal Dakar	1.00

(a) 1973-77 cont	SRm	(SRv)	Highest SRm		Lowest SRm	
Thyroid	0.43	(0.02)	Poland Nowy Sacz Rural	1.11	Poland Warsaw Rural	0.14
			UK Scotland North	0.82	Spain Zaragoza	0.22
			Slovenia	0.71	India Poona	0.22
Other endocrine glands	1.57	(1.36)	Netherlands Antilles	5.00	Switzerland Neuchatel	0.33
			USA Utah	4.00	Colombia Cali	0.40
			Japan Miyagi Prefacture	4.00	Jamaica Kingston & St Andrew	0.50
Bone	1.57	(0.91)	Netherlands Antilles	8.00	Senegal Dakar	0.67
			Poland Nowy Sacz Rural	4.00	France Doubs	0.73
			Italy Varese	3.75	France Bas-Rhin	0.85
Skin melanoma	1.07	(1.03)	Spain Zaragoza	2.00	UK Scotland North East	0.42
			Colombia Cali	1.73	Jamaica Kingston & St Andrew	0.50
			Germany Hamburg	1.60	Netherlands Antilles	0.50
Other skin cancers	1.54	(0.11)	Spain Navarra	2.34	India Poona	0.89
			Germany Hamburg	2.26	Canada NWT & Yukon	0.98
			Sweden	2.24	Slovakia	0.98
Connective tissue	1.39	(0.55)	Canada NWT & Yukon	5.71	Poland Nowy Sacz	0.40
			UK Scotland East	3.57	France Doubs	0.50
			Switzerland Neuchatel	2.22	Poland Warsaw Rural	0.60
Kidney	4.19	(3.17)	France doubs	11.58	Canada NWT & Yukon	0.74
			Spain Navarra	9.63	Senegal Dakar	1.76
			Spain Zaragoza	8.21	India Poona	2.00
Bladder	2.03	(0.96)	Japan Nagasaki	4.33	Canada NWT & Yukon	0.46
			Australia South	3.32	Senegal Dakar	0.83
			Switzerland Geneva	3.22	Switzerland Neuchatel	0.89
Hodgkin lymphoma	1.73	(0.54)	Australia South	3.50	France Doubs	0.25
			Switzerland Neuchatel	3.42	Hungary County Vas	1.00
			Poland Cieszyn	2.75	China Shanghai	1.10

(a) 1973-77 cont	SRm	(SRv)	Highest SRm		Lowest SRm	
Non-Hodgkin lymphoma	1.56	(0.16)	Poland Nowy Sacz	3.09	Switzerland Neuchatel	0.82
			Senegal Dakar	2.92	Canada NWT & Yukon	0.83
			Spain Zaragoza	2.71	UK Scotland North	0.86
Multiple myeloma	1.46	(0.41)	China Shanghai	6.00	Japan Nagasaki	0.61
			India Poona	3.00	Poland Warsaw Rural	0.62
			Netherlands Antilles	2.82	Hungary Szabolcs-Szatmar	0.71
Leukemia	1.52	(0.17)	Senegal Dakar	3.50	Canada NWT & Yukon	0.63
			Switzerland Neuchatel	3.30	Jamaica Kingston & St Andrew	0.97
			Poland Cieszyn	2.53	India Poona	1.09
Eye	1.38	(0.59)	Canada NF	3.50	UK Scotland North East	0.38
			Spain Navarra	3.00	Switzerland Neuchatel	0.60
			Spain Nagasaki	3.00	USA Washington Seattle	0.63
Brain	1.47	(0.84)	Switzerland Geneva	2.29	Canada NWT & Yukon	0.85
			Jamaica Kingston & St Andrew	2.18	Poland Cieszyn	0.90
				1.89	Netherlands Antilles	0.96

(b) 1988-1992 – Cancer Incidence in Five Continents Volume VII

	Overall SR		Populations with highest & lowest SRm			
	SRm	(SRv)	Highest SRm		Lowest SRm	
Lip	7.40	(43.22)	Canada PEI	38.00	Thailand Khon Kaen	0.11
			Spain Navara	32.50	Vietnam Hanoi	0.50
			Spain Zaragosa	31.50	Thailand Chiang Mai	0.50
Tongue	4.17	11.05	Slovakia	17.00	China Qidong	0.50
			Brazil Goiania	14.50	Zimbabwe Harare	0.51
			Poland Cracow	14.00	Kuwait	0.67
Salivary glands	1.78	1.08	France Tarn	7.00	Malta	0.20
			Ireland, Southern	6.00	Switzerland Valais	0.25
			India Barshi	5.00	Uganda Kyadondo	0.37
Mouth	3.65	9.88	Switzerland Graubunden	22.00	India Bangalore	0.31
			Slovakia	13.50	Italy Macereta	0.50

(b) 1988-92 cont…			Spain Zaragosa	13.50	Thailand Khon Kaen	0.57
Pharynx	6.12	23.47	Spain Tarragona	25.00	Canada Yukon	0.79
			France Calvados	23.42	Kuwait	1.25
			France Somme	20.14	Zimbabwe Harare	1.37
Larynx	15.26	433.39	Spain Zaragosa	171.00	Uganda Kyadondo	0.44
			Spain Granada	137.00	China Tianjin	1.94
			Spain Murcia	79.50	Canada NWT	2.27
Lung	5.37	13.13	Spain Zaragosa	17.78	Thailand Chiang Mai	1.19
			Spain Granada	16.78	Iceland	1.24
			Spain Basque	15.62	Canada NWT	1.38
Oesophagus	5.42	21.49	Spain Albacete	31.00	Canada NWT	0.60
			Spain Tarragona	28.50	Kuwait	0.68
			Korea Kangwha	20.40	India Bangalore	1.04
Nose & sinuses	2.58	3.21	Italy Ventian region	10.00	Italy Modena	0.50
			France Doubs	10.00	India Karunagappally	0.50
			Italy Florence	9.00	Malta	10.50
Stomach	2.22	0.22	Algeria Setif	4.11	Zimbabwe Harare	0.76
			Canada NWT	4.00	Kuwait	0.85
			UK East Anglia	3.15	India Barshi	0.89
Small intestine	1.66	1.24	Spain Albacete	10.00	Zimbabwe Harare	0.29
			Switzerland Graubunden	6.00	French Polynesia	0.36
			Spain Murcia	4.50	Spain Navarra	0.50
Colon	1.31	0.07	India Trivandrum	2.40	Canada NWT	0.59
			Zimbabwe Harare	2.21	Brazil Goiana	0.64
			French Polynesia	1.98	Algeria Setif	0.67
Rectum & anus	1.67	0.22	India Karunagappally	5.33	Brazil Belem	0.60
			Zimbabwe Harare	3.03	Ecuador Quito	0.85

(b) 1988-92			Mali Bamako	3.00	Brazil Goiana	0.87
Liver	2.96	2.57	France Isere	10.11	Peru Lima	1.00
			France Haut-Rhin	9.50	Ecuador Quito	1.04
			France Calvados	8.87	Colombia Cali	1.22
Gallbladder	0.84	0.12	Vietnam Hanoi	3.00	Brazil Porto Alegre	0.29
			French Polynesia	2.19	India Trivandrum	0.33
			Italy Macerata	1.76	Poland Lower Silesia	0.35
Pancreas	1.64	0.69	India Karunagappally	10.50	Zimbabwe Harare	0.74
			Malta	2.83	Peru Lima	0.83
			Korea Kangwha	2.67	Ecuador Quito	0.86
Thyroid	0.38	0.05	India Barshi	2.50	Spain Albacete	0.05
			Australia Capital Territory	1.05	Mali Bamako	0.05
			Canada NWT	0.75	Switzerland Neuchatel	0.10
Other endocrine glands	1.29	0.66	Estonia	4.00	Iceland	0.33
			Spain Asturias	3.33	Switzerland Neuchatel	0.40
			Canada NB	3.00	France Isere	0.43
Bone	1.51	0.59	Switzerland Geneva	6.50	Spain Albacete	0.33
			Uganda Kyadondo	3.83	Switzerland Neuchatel	0.33
			Italy Latina	3.67	Malta	0.40
Skin melanoma	1.00	0.14	Japan Miyagi	3.00	Mali Bamako	0.33
			India Trivandrum	3.00	Zimbabwe Harare	0.46
			Peru Trujillo	1.82	France Calvados	0.51
Other skin cancers	1.57	0.32	Korea Kangwha	5.00	Spain Albacete	0.55
			France Somme	3.37	Puerto Rico	0.70
			Uganda Kyadondo	3.10	USA Utah	0.89
Connective tissue	1.39	0.62	Canada NWT	9.00	Brazil Goiana	0.54
			Zimbabwe Harare	3.83	Switzerland Valais	0.56

(b) 1988-92			Canada PEI	3.44	Mali Bamako	0.60
Kidney	1.99	0.29	India Barshi	4.00	Uganda Kyadondo	0.50
			French Polynesia	3.67	Mali Bamako	0.68
			Canada PEI	3.46	Argentina Concordia	0.97
Bladder	5.06	4.83	Algeria Setif	18.00	Zimbabwe Harare	1.06
			Spain Albacete	10.94	Thailand Chiang Mai	2.12
			Italy Macerata	10.81	Argentina Concordia	2.33
Hodgkin lymphoma	1.46	0.81	Peru Trujillo	9.00	Japan Saga Prefacture	0.50
			Uganda Kyadondo	4.50	Switzerland Valais	0.62
			Switzerland Graubunden	3.55	Canada NWT	0.68
Non-Hodgkin lymphoma	1.54	0.15	India Karunagappally	4.64	Kuwait	0.77
			Japan Saga Prefacture	2.67	Canada NWT	0.83
			Mali Bmako	2.44	French Polynesia	0.89
Multiple myeloma	1.43	0.29	India Trivandrum	6.50	India Karunagappally	0.43
			China Qidong	2.60	Korea Kangwha	0.50
			Japan Yamagata	2.11	Zimbabwe Harare	0.67
Leukemia	1.49	0.11	Korea Kangwha	3.08	Mali Bamako	0.30
			Switzerland Neuchatel	2.44	Uganda Kyadondo	0.58
			Spain Albacete	2.38	Argentina Concordia	0.94
Eye	1.43	0.48	Italy Trieste	5.00	India Trivandrum	0.50
			Japan Hiroshima	4.00	Italy Romagna	0.56
			Spain Murcia	3.50	Brazil Goiana	0.63
Brain	1.49	0.68	Canada Yukon	10.55	French Polynesia	0.67
			Algeria Setif	2.67	Italy Macereta	0.80
			Argentina Concordia	2.54	Italy Ferrara	0.96

(b) 2003-2007 – Cancer Incidence in Five Continents Volume X

Cancer	Overall SR		Populations with highest & lowest SRm			
	SRm	(SRv)	Highest SRm		Lowest SRm	
Lip	4.56	(12.94)	Australia Northern Territory	23.00	Thailand Khon Khen	0.19
			Italy Sassari Province	18.00	India Karunagappally	0.33
			Iceland	17.00	Thailand Lampang	0.33
Tongue	3.01	(3.00)	Belarus	15.00	Colombia Pasto	0.33
			Portugal Azores	10.20	Canada Yukon	0.56
			Argentina Tierra del Fuego	9.67	China Jiashan	0.60
Salivary glands	1.56	(0.96)	Italy Lombardy Mantova	12.00	Canada PEI	0.17
			France Vendee	8.00	Portugal Azores	0.25
			Switzerland Valais	5.50	Korea Jejudo	0.25
Mouth	2.54	(2.44)	Belarus	12.33	China Jiaxin	0.33
			Ukraine	11.33	Kuwait	0.43
			Latvia	8.33	Canada Yukon	0.44
Pharynx	3.34	(4.98)	Spain Canary Islands	12.00	Libya Benghazi	0.50
			Brazil Cuiaba	11.00	Australia Capital Territory	0.50
			Spain Austurias	10.00	Switzerland Neuchatel	0.67
Larynx	11.53	(104.52)	Spain Cuenca	100.00	Canada NWT	0.54
			China Zhongshan	55.00	India Sikkim	1.50
			Korea Jejudo	52.00	Malawi Blantyre	2.00
Lung	3.28	(4.85)	Turkey Edirne	15.60	India Barshi Paranda & Bhum	0.70
			Belarus	12.08	India Sikkim	0.97
			Tunisia North	11.77	India Mizoram	1.00
Oesophagus	5.01	(11.22)	Spain Cuenca	25.00	Qatar	0.57
			Korea Jejudo	20.75	Colombia Manizales	0.80
			Spain Albacete	30.50	India Sikkim	1.04

(c) 2003-07 cont…	SRm	(SRv)	Highest SRm		Lowest SRm	
Nose & sinuses	2.34	(1.78)	Spain La Rioja	10.00	Uganda Kyadondo	0.22
			Malaysia Penang	9.00	Switzerland St. Gall Appenzel	0.40
			France Loire Atlantique	7.50	India Sikkim	0.44
Stomach	2.15	(0.18)	India Trivandrum	4.77	Canada NWT	0.74
			France Manche	3.87	Malawi Blantyre	0.80
			France Tarn	3.35	Zimbabwe Harare	0.82
Small intestine	1.63	(0.89)	Korea Jejudo	9.00	Argentina Tierra del Fuego	0.28
			Zimbabwe Harare	8.00	Italy Syracuse	0.29
			Italy Nuoro	3.00	Thailand Lampang	0.33
Colon	1.38	(0.07)	India Barshi Paranda & Bhum	3.67	India Sikkim	0.73
			Spain Basque	2.06	Ecuador Cuenca	0.77
			Korea Jejudo	2.00	Chile Biobio	0.81
Rectum & anus	1.74	(0.14)	Argentina Tierra del Fuego	3.80	Ecuador Cuenca	0.58
			India Sikkim	3.50	Uganda Kyandondo	0.83
			India Barshi Paranda & Bhum	2.67	Malawi Blantyre	0.83
Liver	3.04	(1.64)	Argentina Tierra del Fuego	12.17	Ecuador Cuenca	0.69
			Italy Sondrio	8.78	Cuba Villa Clara	1.09
			France Vendee	8.47	Algeria Setif	1.19
Gallbladder	0.99	(0.10)	France Calvados	2.40	India Sikkim	0.26
			Canada NWT	2.37	Qatar	0.28
			Switzerland Graubunden & Glarus	2.13	Algeria Setif	0.36
Pancreas	1.44	(0.09)	Iran Golestan	2.80	Malawi Blantyre	0.25
			China Yanting	2.37	Brazil Aracaju	0.60
			Korea Jejudo	2.33	Zimbabwe Harare	0.70
Thyroid	0.33	(0.01)	Canada PEI	1.00	Argentina Tierra del Fuego	0.03
			China Yanting	1.00	Canada Yukon	0.08
			Thailand Bangkok	0.97	Korea Daejeon	0.13

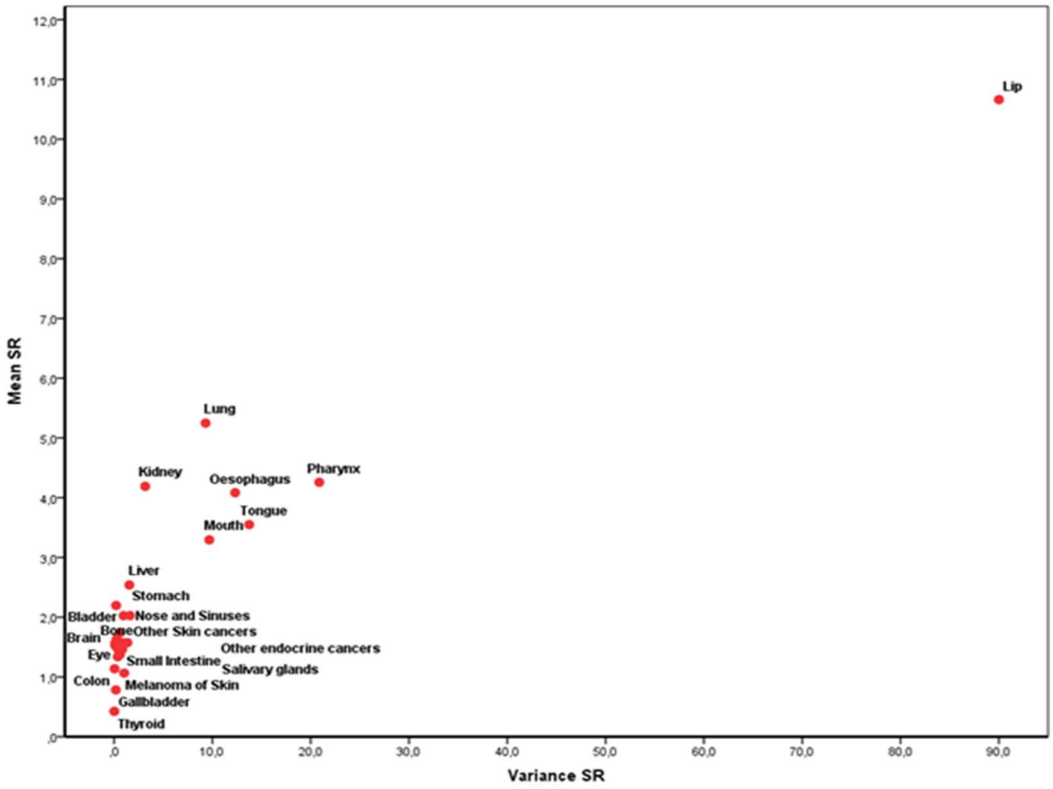
(c) 2003-07 cont….	SRm		Highest SRm		Lowest SRm	
Other endocrine glands	1.65	(1.12)	Italy Parma	7.00	Spain Albacete	0.25
			Italy Nuoro	6.00	Italy Naples	0.33
			Australia Tasmania	6.00	Portugal Azores	0.33
Bone	1.43	(0.41)	Italy Sassari	5.33	Turkey Edrine	0.29
			Switzerland Valais	5.00	Italy Nuoro	0.33
			Bahrain	4.00	Libya Benghazi	0.50
Skin melanoma	1.09	(0.12)	India Bhopal	4.00	China Macao	0.25
			China Yanting	3.00	Bahrain	0.33
			Thailand Bangkok	2.00	Malawi Blantyre	0.37
Other skin cancers	1.41	(0.14)	France Martinique	4.00	Spain Cuenca	0.22
			USA Wyoming	3.00	UK wales	0.50
			Switzerland Tiando	2.50	Canada Saskatchwan	0.50
Connective tissue	1.38	(0.17)	China Macao	4.50	China Cixian	0.33
			Bahrain	3.17	Spain Cuenca	0.47
			China Qideng	3.00	China Jiashan	0.50
Kidney	2.09	(0.41)	India Karunagappally	8.00	China Yanting	0.50
			Canada Yukon	5.88	Uganda Kyadondo	0.65
			India Trivandrum	5.00	South Africa	0.75
Bladder	4.66	(3.58)	India Karunagappally	15.00	Canada NWT	1.02
			Turkey Trabzon	13.83	India Sikkim	1.33
			Italy Nuoro	13.71	Uganda Kyandondo	1.44
Hodgkin lymphoma	1.35	(0.41)	India Karunagappally	8.00	China Nangang Harbin	0.33
			Malaysia Penang	5.33	China Jiashan	0.50
			Korea Gwangju	4.00	India Mizoram	0.50
Non-Hodgkin lymphoma	1.45	(0.09)	China Yanting	4.07	India Sikkim	0.45
			Canada NWT	3.59	Ecuador Cuenca	0.84
			India Dindigul Ambillikai	2.50	Chile Antofagasta	0.87

(c) 2003-07 cont…	SRm	(SRv)	Highest SRm		Lowest SRm	
Multiple myeloma	1.42	(0.23)	Canada Yukon	7.00	India Mizoram	0.33
			South Africa	3.50	Colombia Pasto	0.43
			Qatar	3.00	Zimbabwe Harare	0.45
Leukemia	1.52	(0.90)	Malaysia Penang	10.00	Spain Cuenca	0.17
			Zimbabwe Harare	7.00	France Somme	0.20
			Brazil Cuiaba	4.33	Chile Biobio	0.25
Eye	1.51	(1.55)	Switzerland Neuchatel	14.00	Korea Gwangju	0.33
			China Yangcheng	6.50	Colombia Pasto	0.33
			Spain Granada	6.00	Jamaica Kingston & St. Andrew	0.40
Brain	1.37	(0.07)	China Yanting	2.23	Malawi Blantyre	0.33
			Uganda Kyandondo	2.11	Zimbabwe Harare	0.50
			Italy Genova	2.06	Jamaica Kingston & St. Andrew	0.58

**Cancers are listed in ascending order of ICD-O-3 codes*

Appendix 5: Magnitude of sex ratios (un-transformed versions) of 30 cancer types plotted against their variances in (a) 1973-1977, (b) 1988-92, and (c) 2003-07.

(a) 1973-77



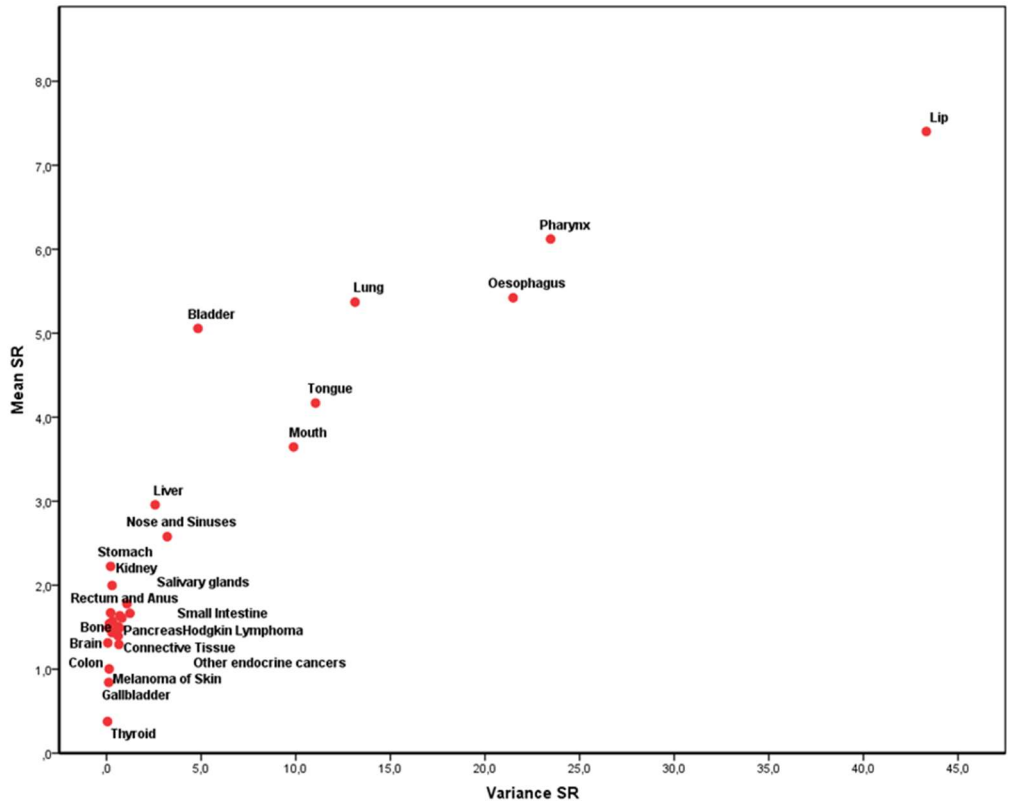
Cancer types categorized according to three levels of logs of sex ratio magnitude (Log SRm) and variances (Log SRv) based on tertiles.

		SRv		
		Low	***Medium***	***High***
SRm	***Low***	Salivary glands, Gallbladder, Thyroid, Colon	Small Intestine, Connective tissues, Eye, Skin Melanoma, Brain, Multiple Myeloma	
	Medium	Rectum & Anus, Pancreas, Non-Hodgkin Lymphoma, Leukemia, Other skin cancers	Bone, Hodgkin Lymphoma, Other endocrine glands, Bladder	Nose & Sinuses
	High	Stomach		Tongue, Mouth, Lip, Pharynx, Oesophagus, Liver, Larynx, Lung, Kidney

SRm (Low): ≤ 1.500; SRm (Medium): 1.511 to 2.030; SRm (High): > 2.030
SRv (Low): ≤ 0.38); SRv (Medium): 0.39 to 1.36; SRv (High): > 1.36

Appendix 5 cont….

(b) 1988-92

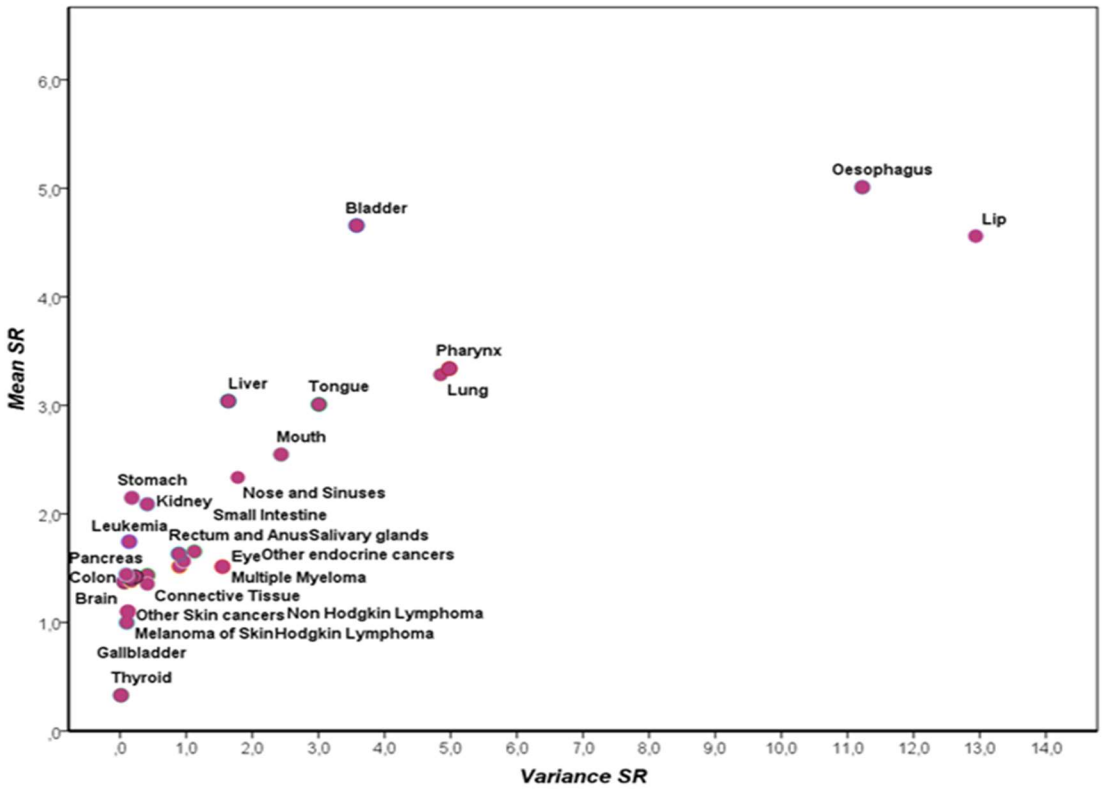


Cancer types categorized according to three levels of logs of sex ratio magnitude (Log SRm) and variances (Log SRv) based on tertiles

		Log SRv		
		Low	**Medium**	**High**
Log SRm	**Low**	Colon, Gallbladder, Skin melanoma, Thyroid, Multiple myeloma	Bone, Connective tissues, Eye, Brain, Other endocrine glands	
	Medium	Stomach, Rectum & anus, Kidney, Non-Hodgkin Lymphoma, Leukemia	Salivary glands, Small Intestine, Pancreas, Hodgkin Lymphoma, Other skin cancers	
	High			Lip, Tongue, Pharynx, Oesophagus, Liver, Nose & Sinuses, Larynx, Bladder

SRm (Low): ≤ 1.49; SRm (Medium): 1.50 to 2.56; SRm (High): > 2.56
SRv (Low): ≤ 0.30); SRv (Medium): 0.31 to 1.24; SRv (High): > 1.24

(c) 2003-07



Cancer types categorized according to three levels of logs of sex ratio magnitude (Log SRm) and variances (Log SRv) based on tertiles

		Var log SR		
		Low	**Medium**	**High**
Mag log SR	**Low**	Colon, Gallbladder, Skin melanoma, Connective tissue, Brain, Thyroid	Bone, Hodgkin Lymphoma, Multiple Myeloma, Other skin cancers	
	Medium	Other endocrine glands, Rectum & Anus, Pancreas, Non-Hodgkin Lymphoma,	Salivary glands, Stomach Small Intestine, Kidney, Eye, Leukemia,	
	High			Lip, Tongue, Mouth, Pharynx, Oesophagus, Liver,Nose & Sinuses, Larynx, Lung, Bladder

SRm (Low): ≤ 1.43; SRm (Medium): 1.44 to 2.15; SRm (High): > 2.15
SRv (Low): ≤ 0.17); SRv (Medium): 0.18 to 1.55; SRv (High): > 1.55

Appendix 6: Regression models in three scenarios (A, B & C) with varying number of registry-populations (N) for 21 types of cancers (A 6-a to A 6-u): fixed-effects models (a and b); mixed effects model with random intercepts (c); and mixed effects model with random intercepts and slopes (d).

- 21 types of cancers (A 6-a to A 6-u) starting on next 20 pages. Formal notations in all four models are explained in appendix 2. In the footnotes of tables A 6-a to A 6-u, basic notations are used that carries same meanings as formal notations shown in appendix 2.
- Three scenario A, B, and C are explained in the text in section 7.6.1 of Chapter 7.
- Registries are used as surrogates of regions (for the intra- and inter-regional variations).
- Model d (RI and RS models) from the appendix 6 (A 6-a to A 6-u) are used in Tables 8-12 to 8-14 of Chapter 8.

A 6-a: Cancer of oral cavity and pharynx : regression models in three scenarios (A, B & C) with varying number of registry-populations (N): fixed-effects models (a and b); mixed effects model with random intercepts (c); and mixed effects model with random intercepts and slopes (d).

Models	Errors & AIC	βo	β_1	*SE* β_1	(95% CI)	intra- regional variance $(\delta^k)^2$	Between-country variance $(\tau^k)^2$ or τ_1^2	Variation in slopes τ_2^2	Cov $(u_{1i}, u_{2j})\tau_{12}$
(A) 1953-2007 (N= 3 to 113)†									
a. Fixed-Effects	ε 663530	15.27	-0.063	0.026	(-0.115,-0.011)				
b. Fixed Effects	ε 393815	10.46	-0.154	0.023	(-0.200,-0.108)				
c. Random Intercepts	AIC 22735	19.26	-0.149	0.023	(-0.194,-0.103)	142.70*	τ^2 101.42*		
d. Random Intercepts & Slopes	AIC 22566	21.27	-0.193	0.0535	(-0.299,-0.086)	130.50*	τ_1^2 838.44*	τ_2^2 0.22*	τ_{12} -13.31*
(B) 1983-2007 (N = 76 to 113) †									
a. Fixed-Effects	ε 568028	19.94	-0.167	0.046	(-0.258,-0.076)				
b. Fixed Effects	ε 333546	13.37	-0.209	0.037	(-0.283,-0.135)				
c. Random Intercepts	AIC 18564	21.82	-0.205	0.037	(-0.279,-0.132)	150.79*	τ^2 99.58*		
d. Random Intercepts & Slopes	AIC 18350	22.08	-0.211	0.0706	(-0.350,-0.071)	132.51*	τ_1^2 1301.22*	τ_2^2 0.39*	τ_{12} -22.36*
(C) 1983-2007 (N = 76) ‡									
a. Fixed-Effects	ε 410347	19.86	-0.185	0.051	(-0.286,-0.085)				
b. Fixed Effects	ε 249146	12.66	-0.192	0.041	(-0.272,-0.112)				
c. Random Intercepts	AIC 13825	20.15	-0.192	0.041	(-0.272,-0.112)	150.01*	τ^2 90.15*		
d. Random Intercepts & Slopes	AIC 13673	20.20	-0.193	0.079	(-0.349,-0.036)	132.01*	τ_1^2 1096.01*	τ_2^2 0.35*	τ_{12} -19.19*

a. Sex ratio $_i = \beta o + \beta_1$ (Time) + ε_i $\varepsilon i \sim N(0, \delta^2)$ * p value < 0.005

b. Sex ratio $_i = \beta o + \beta_1$ (Time) + β_2 (country1) + β_3 (country2) ….. + β_n (country n) + ε_i $\varepsilon i \sim N(0, \delta^2)$

c. Sex ratio $_{ij} = \beta o + \beta_1$ (Time) + $\varepsilon_{ij} + u_j$ $\varepsilon ij \sim N(0, \delta^2)$; $u_j \sim N(0, \tau^2)$ for country j

d. Sex ratio $_{ij} = \beta o + \beta_1$ (Time) + $\varepsilon_{ij} + u_{1j} + u_{2j}$ Time $\varepsilon ij \sim N(0, \delta^2)$; $u_{1j} \sim N(0, \tau_1^2)$ for country j; $u_{2j} \sim N(0, \tau_2^2)$ for country j; Cov $(u_{1i}, u_{2j}) = \tau_{12}$

† Unbalanced data (No. of countries different in different year) ‡ Balanced data (No. of countries same in each year) AIC gives an indication of which model to choose (lower is better)

A 6-b: Cancer of Lung : regression models in three scenarios (A, B & C) with varying number of registry-populations (N): fixed-effects models (a and b); mixed effects model with random intercepts (c); and mixed effects model with random intercepts and slopes (d).

Models	Errors & AIC	βo	β_1	*SE* β_1	(95% CI)	intra- regional variance δ^2 (ε_{ij})	inter-regional variance τ^2 or τ_1^2	Variation in slopes τ_2^2	Cov (u_{1i}, u_{2j}) τ_{12}
(A) 1953-2007 (N= 3 to 113)†									
a. Fixed-Effects	ε 47211	8.31	-0.096	0.006	(-0.108, -0.083)				
b. Fixed Effects	ε 15644	6.83	-0.134	0.004	(-0.143, -0.126)				
c. Random Intercepts	AIC 14593	10.09	-0.133	0.004	(-0.142, -0.125)	5.14*	τ^2 10.69*		
d. Random Intercepts & Slopes	AIC 13586	11.89	-0.173	0.018	(-0.208, -0.137)	3.45*	τ_1^2 114.64*	τ_2^2 0.03*	τ_{12} -1.89*
(B) 1983-2007 (N = 76 to 113) †									
a. Fixed-Effects	ε 34978	9.84	-0.130	0.010	(-0.151, -0.110)				
b. Fixed Effects	ε 10888	7.89	-0.152	0.006	(-0.164, -0.140)				
c. Random Intercepts	AIC 11533	10.87	-0.151	0.006	(-0.163, -0.139)	4.45*	τ^2 9.75*		
d. Random Intercepts & Slopes	AIC 10963	11.71	-0.169	0.017	(-0.203, -0.134)	3.36*	τ_1^2 113.44*	τ_2^2 0.03*	τ_{12} -1.84*
(C) 1983-2007 (N = 76) ‡									
a. Fixed-Effects	ε 19545	9.99	-0.142	0.010	(-0.162, -0.121)				
b. Fixed Effects	ε 4607	7.45	-0.141	0.005	(-0.152, -0.131)				
c. Random Intercepts	AIC 7341	9.98	-0.141	0.005	(-0.152, -0.131)	2.59*	τ^2 7.97*		
d. Random Intercepts & Slopes	AIC 6573	9.98	-0.141	0.017	(-0.174, -0.109)	1.57*	τ_1^2 69.42*	τ_2^2 0.019*	τ_{12} -1.129*

a. Sex ratio $_i = \beta o + \beta_1$ (Time) + ε_i $\varepsilon i \sim N(0, \delta^2)$ * p value < 0.005

b. Sex ratio $_i = \beta o + \beta_1$ (Time) + β_2 (country1) + β_3 (country2) ….. + β_n (country n) + ε_i $\varepsilon i \sim N(0, \delta^2)$

c. Sex ratio $_{ij} = \beta o + \beta_1$ (Time) + $\varepsilon_{ij} + u_j$ $\varepsilon ij \sim N(0, \delta^2)$; $u_j \sim N(0, \tau^2)$ for country j

d. Sex ratio $_{ij} = \beta o + \beta_1$ (Time) + $\varepsilon_{ij} + u_{1j} + u_{2j}$ Time $\varepsilon ij \sim N(0, \delta^2)$; $u_{1j} \sim N(0, \tau_1^2)$ for country j; $u_{2j} \sim N(0, \tau_2^2)$ for country j; Cov (u_{1i}, u_{2j}) = τ_{12}

† Unbalanced data (No. of countries different in different year) ‡ Balanced data (No. of countries same in each year) AIC gives an indication of which model to choose (lower is better)

A 6-c: Esophageal cancer : regression models in three scenarios (A, B & C) with varying number of registry-populations (N): fixed-effects models (a and b); mixed effects model with random intercepts (c); and mixed effects model with random intercepts and slopes (d).

Models	Errors & AIC	βo	β_1	*SE* β_1	(95% CI)	intra- regional variance δ^2 (ε_{ij})	inter-regional variance τ^2 or τ_1^2	Variation in slopes τ_2^2	Cov (u_{1i}, u_{2j}) τ_{12}
(A) 1953-2007 (N= 3 to 113)†									
a. Fixed-Effects	ε 152267	4.70	0.024	0,012	(0.001, 0.047)				
b. Fixed Effects	ε 98196	4.00	-0.015	0.011	(0.001, 0.007)				
c. Random Intercepts	AIC 19736	6.37	-0.012	0.011	(-0.033, 0.009)	33.31*	τ^2 18.06*		
d. Random Intercepts & Slopes	AIC 19559	8.15	-0.052	0.023	(-0.099,-0.005)	30.45*	τ_1^2 146.78*	τ_2^2 0.04*	τ_{12} -2.40*
(B) 1983-2007 (N = 76 to 113) †									
a. Fixed-Effects	ε 100425	8.06	-0.053	0.018	(-0.089,-0.016)				
b. Fixed Effects	ε 58006	6.25	-0.061	0.015	(-0.090,-0.032)				
c. Random Intercepts	AIC 15214	8.48	-0.060	0.015	(-0.089,-0.031)	24.67*	τ^2 16.75*		
d. Random Intercepts & Slopes	AIC 15105	8.73	-0.066	0.022	(-0.109,-0.022)	23.26*	τ_1^2 129.48*	τ_2^2 0.030*	τ_{12} -1.95*
(C) 1983-2007 (N = 76) ‡									
a. Fixed-Effects	ε 74133	8.02	-0.056	0.021	(-0.097,-0.014)				
b. Fixed Effects	ε 41501	5.96	-0.054	0.016	(-0.086,-0.023)				
c. Random Intercepts	AIC 11059	7.92	-0.054	0.016	(-0.086,-0.023)	24.03*	τ^2 17.06*		
d. Random Intercepts & Slopes	AIC 10953	7.83	-0.052	0.026	(-0.104,-0.0001)	22.24*	τ_1^2 132.53*	τ_2^2 0.033*	τ_{12} -2.06*

a. Sex ratio $_i = \beta o + \beta_1$ (Time) + ε_i $\varepsilon i \sim N(0, \delta^2)$ * p value < 0.005

b. Sex ratio $_i = \beta o + \beta_1$ (Time) + β_2 (country1) + β_3 (country2) ….. + β_n (country n) + ε_i $\varepsilon i \sim N(0, \delta^2)$

c. Sex ratio $_{ij} = \beta o + \beta_1$ (Time) + $\varepsilon_{ij} + u_j$ $\varepsilon ij \sim N(0, \delta^2)$; $u_j \sim N(0, \tau^2)$ for country j

d. Sex ratio $_{ij} = \beta o + \beta_1$ (Time) + $\varepsilon_{ij} + u_{1j} + u_{2j}$ Time $\varepsilon ij \sim N(0, \delta^2)$; $u_{1j} \sim N(0, \tau_1^2)$ for country j; $u_{2j} \sim N(0, \tau_2^2)$ for country j; Cov (u_{1i}, u_{2j}) = τ_{12}

† Unbalanced data (No. of countries different in different year) ‡ Balanced data (No. of countries same in each year) AIC gives an indication of which model to choose (lower is better)

A 6-d: Cancer of oral cavity and pharynx : regression models in three scenarios (A, B & C) with varying number of registry-populations (N): fixed-effects models (a and b); mixed effects model with random intercepts (c); and mixed effects model with random intercepts and slopes (d).

Models	Errors & AIC	βo	β_1	*SE* β_1	(95% CI)	intra- regional variance δ^2 (ε_{ij})	inter-regional variance τ^2 or τ_1^2	Variation in slopes τ_2^2	Cov (u_{1i}, u_{2j}) τ_{12}
(A) 1953-2007 (N= 113)									
Fixed-Effects [a]	ε 29837	5.16	- 0.028	0.005	(-0.038,-0.019)				
Fixed Effects [b]	ε 13049	5.23	-0.048	0.004	(-0.056,-0.040)				
Random Intercepts [c]	AIC 13936	5.99	-0.047	0.004	(-0.054,-0.039)	4.31*	τ^2 5.15*		
Random Intercepts & Slopes [d]	AIC 13451	6.88	-0.067	0.012	(-0.090,-0.042)	3.52*	τ_1^2 46.58*	τ_2^2 0.013*	τ_{12} -0.77*
(B) 1983-2007 (N = 76-113) [e]									
Fixed-Effects [a]	ε 25954	6.87	-0.067	0.008	(-0.083,-0.052)				
Fixed Effects [b]	ε 11417	5.32	-0.075	0.006	(-0.086,-0.064)				
Random Intercepts [c]	AIC 12084	7.22	-0.074	0.006	(-0.085,-0.064)	4.39*	τ^2 5.05*		
Random Intercepts & Slopes [d]	AIC 11603	7.44	-0.079	0.013	(-0.105,-0.053)	3.53*	τ_1^2 55.44*	τ_2^2 0.016*	τ_{12} -0.93*
(C) 1983-2007 (N = 76) [f]									
Fixed-Effects [a]	ε 20983	7.57	-0.084	0.009	(-0.102,-0.065)				
Fixed Effects [b]	ε 9357	6.87	-0.087	0.006	(-0.100,-0.075)				
Random Intercepts [c]	AIC 9707	7.73	-0.087	0.006	(-0.099,-0.074)	4.51*	τ^2 5.16*		
Random Intercepts & Slopes [d]	AIC 9264	7.80	-0.089	0.016	(-0.120,-0.057)	3.50*	τ_1^2 65.22*	τ_2^2 0.019*	τ_{12} -1.11*

a. Sex ratio $_i = \beta o + \beta_1$ (Time) + ε_i $\varepsilon i \sim N(0, \delta^2)$ * p value < 0.005

b. Sex ratio $_i = \beta o + \beta_1$ (Time) + β_2 (country1) + β_3 (country2) ….. + β_n (country n) + ε_i $\varepsilon i \sim N(0, \delta^2)$

c. Sex ratio $_{ij} = \beta o + \beta_1$ (Time) + $\varepsilon_{ij} + u_j$ $\varepsilon ij \sim N(0, \delta^2)$; $u_j \sim N(0, \tau^2)$ for country j

d. Sex ratio $_{ij} = \beta o + \beta_1$ (Time) + $\varepsilon_{ij} + u_{1j} + u_{2j}$ Time $\varepsilon ij \sim N(0, \delta^2)$; $u_{1j} \sim N(0, \tau_1^2)$ for country j; $u_{2j} \sim N(0, \tau_2^2)$ for country j; Cov (u_{1i}, u_{2j}) = τ_{12}

† Unbalanced data (No. of countries different in different year) ‡ Balanced data (No. of countries same in each year) AIC gives an indication of which model to choose (lower is better)

A 6-e: Cancer of bladder: regression models in three scenarios (A, B & C) with varying number of registry-populations (N): fixed-effects models (a and b); mixed effects model with random intercepts (c); and mixed effects model with random intercepts and slopes (d).

Models	Errors & AIC	βo	β_1	*SE* β_1	(95% CI)	intra- regional variance δ^2 (ε_{ij})	inter-regional variance τ^2 or τ_1^2	Variation in slopes τ_2^2	Cov (u_{1i}, u_{2j}) τ_{12}
(A) 1953-2007 (N= 3 to 113)†									
a. Fixed-Effects	ε 17710	4.06	0.016	0.004	(0.009, 0.024)				
b. Fixed Effects	ε 10247	4.01	-0.002	0.003	(-0.009, 0.005)				
c. Random Intercepts	AIC 13070	4.88	-0.001	0.003	(-0.008, 0.006)	3.39*	τ^2 2.56*		
d. Random Intercepts & Slopes	AIC 13022	5.04	-0.005	0.006	(-0.173, 0.008)	3.23*	τ_1^2 7.74*	τ_2^2 0.002*	τ_{12} -0.109*
(B) 1983-2007 (N = 76 to 113) †									
a. Fixed-Effects	ε 15761	4.83	-0.001	0.007	(-0.015, 0.013)				
b. Fixed Effects	ε 8980	4.41	-0.010	0.006	(-0.021, 0.001)				
c. Random Intercepts	AIC 10869	5.24	-0.009	0.006	(-0.020, 0.002)	3.69*	τ^2 2.57*		
d. Random Intercepts & Slopes	AIC 10856	5.28	-0.101	0.007	(-0.024, 0.004)	3.59*	τ_1^2 8.33*	τ_2^2 0.002*	τ_{12} -0.114*
(C) 1983-2007 (N = 76) ‡									
a. Fixed-Effects	ε 10637	5.09	-0.011	0.008	(-0.027, 0.004)				
b. Fixed Effects	ε 6047	4.47	-0.012	0.006	(-0.023, 0.001)				
c. Random Intercepts	AIC 7724	5.09	-0.012	0.006	(-0.023, 0.001)	3.41*	τ^2 2.35*		
d. Random Intercepts & Slopes	AIC 7719	5.09	-0.012	0.007	(-0.025, 0.003)	3.36*	τ_1^2 6.22*	τ_2^2 0.001	τ_{12} -0.068*

a. Sex ratio $_i = \beta o + \beta_1$ (Time) + ε_i $\varepsilon i \sim N(0, \delta^2)$ * p value < 0.005

b. Sex ratio $_i = \beta o + \beta_1$ (Time) + β_2 (country1) + β_3 (country2) ….. + β_n (country n) + ε_i $\varepsilon i \sim N(0, \delta^2)$

c. Sex ratio $_{ij} = \beta o + \beta_1$ (Time) + $\varepsilon_{ij} + u_j$ $\varepsilon ij \sim N(0, \delta^2)$; $u_j \sim N(0, \tau^2)$ for country j

d. Sex ratio $_{ij} = \beta o + \beta_1$ (Time) + $\varepsilon_{ij} + u_{1j} + u_{2j}$ Time $\varepsilon ij \sim N(0, \delta^2)$; $u_{1j} \sim N(0, \tau_1^2)$ for country j; $u_{2j} \sim N(0, \tau_2^2)$ for country j; Cov (u_{1i}, u_{2j}) = τ_{12}

† Unbalanced data (No. of countries different in different year) ‡ Balanced data (No. of countries same in each year) AIC gives an indication of which model to choose (lower is better)

A 6-f: Cancer of liver : regression models in three scenarios (A, B & C) with varying number of registry-populations (N): fixed-effects models (a and b); mixed effects model with random intercepts (c); and mixed effects model with random intercepts and slopes (d).

Models	Errors & AIC	βo	β_1	*SE* β_1	(95% CI)	intra- regional variance δ^2 (ε_{ij})	inter-regional variance τ^2 or τ_1^2	Variation in slopes τ_2^2	Cov (u_{1i}, u_{2j}) τ_{12}
(A) 1953-2007 (N= 3 to 113)†									
a. Fixed-Effects	ε 20908	2.51	0.002	0.004	(0.100, 0.027)				
b. Fixed Effects	ε 13980	2.38	0.010	0.004	(-0.001, 0.015)				
c. Random Intercepts	AIC 13968	2.99	0.01	0.004	(-0.0003, 0.015)	4.64*	τ^2 2.09*		
d. Random Intercepts & Slopes	AIC 13928	3.24	0.02	0.006	(-0.001, 0.013)	4.54*	τ_1^2 8.13*	τ_2^2 0.001*	τ_{12} -0.10*
(B) 1983-2007 (N = 76 to 113) †									
a. Fixed-Effects	ε 17277	3.73	-0.01	0.01	(-0.023, 0.006)				
b. Fixed Effects	ε 11654	2.88	-0.01	0.01	(-0.020, 0.005)				
c. Random Intercepts	AIC 11462	3.69	-0.01	0.01	(-0.020, 0.005)	4.80*	τ^2 1.96*		
d. Random Intercepts & Slopes	AIC 11451	3.67	-0.01	0.01	(-0.021, 0.006)	4.80*	τ_1^2 6.14*	τ_2^2 0.0004	τ_{12} -0.05*
(C) 1983-2007 (N = 76) ‡									
a. Fixed-Effects	ε 13756	3.78	-0.008	0.01	(-0.026, 0.009)				
b. Fixed Effects	ε 9478	2.96	-0.009	0.01	(-0.024, 0.005)				
c. Random Intercepts	AIC 8473	3.83	-0.009	0.01	(-0.024, 0.005)	5.39*	τ^2 2.13*		
d. Random Intercepts & Slopes	AIC 8469	3.83	-0.009	0.01	(-0.025, 0.006)	5.40*	τ_1^2 5.74*	τ_2^2 0.0003	τ_{12} -0.04

a. Sex ratio $_i = \beta o + \beta_1 (\text{Time}) + \varepsilon_i$ $\varepsilon i \sim N(0, \delta^2)$ * p value < 0.005

b. Sex ratio $_i = \beta o + \beta_1 (\text{Time}) + \beta_2 (\text{country1}) + \beta_3 (\text{country2}) \ldots\ldots + \beta_n (\text{country n}) + \varepsilon_i$ $\varepsilon i \sim N(0, \delta^2)$

c. Sex ratio $_{ij} = \beta o + \beta_1 (\text{Time}) + \varepsilon_{ij} + u_j$ $\varepsilon ij \sim N(0, \delta^2)$; $u_j \sim N(0, \tau^2)$ for country j

d. Sex ratio $_{ij} = \beta o + \beta_1 (\text{Time}) + \varepsilon_{ij} + u_{1j} + u_{2j}$ Time $\varepsilon ij \sim N(0, \delta^2)$; $u_{1j} \sim N(0, \tau_1^2)$ for country j; $u_{2j} \sim N(0, \tau_2^2)$ for country j; Cov (u_{1i}, u_{2j}) = τ_{12}

† Unbalanced data (No. of countries different in different year) ‡ Balanced data (No. of countries same in each year) AIC gives an indication of which model to choose (lower is better)

A 6-g: Hodgkin lymphoma: regression models in three scenarios (A, B & C) with varying number of registry-populations (N): fixed-effects models (a and b); mixed effects model with random intercepts (c); and mixed effects model with random intercepts and slopes (d).

Models	Errors & AIC	βo	β_1	*SE* β_1	(95% CI)	intra- regional variance δ^2 (ε_{ij})	inter-regional variance τ^2 or τ_1^2	Variation in slopes τ_2^2	Cov (u_{1i}, u_{2j}) τ_{12}
(A) 1953-2007 (N= 3 to 113)†									
a. Fixed-Effects	ε 5371	2.22	-0.015	0.002	(-0.019,-0.011)				
b. Fixed Effects	ε 4773	2.05	-0.019	0.002	(-0.024,-0.014)				
c. Random Intercepts	AIC 10273	2.35	-0.018	0.002	(-0.022,-0.013)	1.63*	τ^2 0.15*		
d. Random Intercepts & Slopes	AIC 10257	2.47	-0.021	0.003	(-0.027,-0.014)	1.61*	τ_1^2 1.05*	τ_2^2 0.0003	τ_{12} -0.02
(B) 1983-2007 (N = 76 to 113) †									
a. Fixed-Effects	ε 4589	2.22	-0.023	0.004	(-0.031,-0.015)				
b. Fixed Effects	ε 4083	2.33	-0.025	0.004	(-0.033,-0.017)				
c. Random Intercepts	AIC 8470	2.63	-0.024	0.004	(-0.032,-0.017)	1.74*	τ^2 0.13*		
d. Random Intercepts & Slopes	AIC 8445	2.63	-0.024	0.005	(-0.034,-0.014)	1.69*	τ_1^2 2.73*	τ_2^2 0.001*	τ_{12} -0.05*
(C) 1983-2007 (N = 76) ‡									
a. Fixed-Effects	ε 3105	2.76	-0.028	0.004	(-0.036,-0.019)				
b. Fixed Effects	ε 2808	2.44	-0.028	0.004	(-0.036,-0.020)				
c. Random Intercepts	AIC 6085	2.76	-0.028	0.004	(-0.036,-0.019)	1.62*	τ^2 0.09*		
d. Random Intercepts & Slopes	AIC 6046	2.74	-0.027	0.005	(-0.038,-0.016)	1.58*	τ_1^2 2.62*	τ_2^2 0.001*	τ_{12} -0.05*

a. Sex ratio $_i = \beta o + \beta_1$ (Time) + ε_i $\varepsilon i \sim N(0, \delta^2)$ * p value < 0.005

b. Sex ratio $_i = \beta o + \beta_1$ (Time) + β_2 (country1) + β_3 (country2) ..... + β_n (country n) + ε_i $\varepsilon i \sim N(0, \delta^2)$

c. Sex ratio $_{ij} = \beta o + \beta_1$ (Time) + $\varepsilon_{ij} + u_j$ $\varepsilon ij \sim N(0, \delta^2)$; $u_j \sim N(0, \tau^2)$ for country j

d. Sex ratio $_{ij} = \beta o + \beta_1$ (Time) + $\varepsilon_{ij} + u_{1j} + u_{2j}$ Time $\varepsilon ij \sim N(0, \delta^2)$; $u_{1j} \sim N(0, \tau_1^2)$ for country j; $u_{2j} \sim N(0, \tau_2^2)$ for country j; Cov (u_{1i}, u_{2j}) = τ_{12}

† Unbalanced data (No. of countries different in different year) ‡ Balanced data (No. of countries same in each year) AIC gives an indication of which model to choose (lower is better)

A 6-h: Cancer of Stomach : regression models in three scenarios (A, B & C) with varying number of registry-populations (N): fixed-effects models (a and b); mixed effects model with random intercepts (c); and mixed effects model with random intercepts and slopes (d).

Models	Errors & AIC	βo	β_1	*SE* β_1	(95% CI)	intra- regional variance δ^2 (ε_{ij})	inter-regional variance τ^2 or τ_1^2	Variation in slopes τ_2^2	Cov (u_{1i}, u_{2j}) τ_{12}
(A) 1953-2007 (N= 3 to 113)†									
a. Fixed-Effects	ε 1417	2.09	0.004	0.001	(0.002, 0.006)				
b. Fixed Effects	ε 1113	2.25	0.003	0.001	(0.001, 0.005)				
c. Random Intercepts	AIC 6018	2.14	0.003	0.001	(0.001, 0.005)	0.37*	τ^2 0.01		
d. Random Intercepts & Slopes	AIC 5999	2.16	0.002	0.001	(-0.001,0.005)	0.36*	τ_1^2 0.10	τ_2^2 0.0001	τ_{12} -0.002
(B) 1983-2007 (N = 76 to 113) †									
a. Fixed-Effects	ε 1259	2.38	-0.003	0.002	(-0.01, 0.001)				
b. Fixed Effects	ε 979	2.43	-0.002	0.002	(-0.01, 0.002)				
c. Random Intercepts	AIC 5132	2.36	-0.002	0.002	(-0.01, 0.001)	0.40*	τ^2 0.09		
d. Random Intercepts & Slopes	AIC 5121	2.36	-0.002	0.002	(-0.01, 0.002)	0.39*	τ_1^2 0.09	τ_2^2 0.0001	τ_{12} -0.002
(C) 1983-2007 (N = 76) ‡									
a. Fixed-Effects	ε 1002	2.37	-0.002	0.002	(-0.01, 0.003)				
b. Fixed Effects	ε 798	2.43	-0.002	0.002	(-0.01, 0.002)				
c. Random Intercepts	AIC 3912	2.38	-0.002	0.002	(-0.01, 0.002)	0.45*	τ^2 0.09*		
d. Random Intercepts & Slopes	AIC 3898	2.37	-0.002	0.003	(-0.01, 0.003)	0.44*	τ_1^2 0.10	τ_2^2 0.001	τ_{12} -0.003

a. Sex ratio $_i = \beta o + \beta_1$ (Time) + ε_i $\varepsilon i \sim N(0, \delta^2)$ * p value < 0.005

b. Sex ratio $_i = \beta o + \beta_1$ (Time) + β_2 (country1) + β_3 (country2) ….. + β_n (country n) + ε_i $\varepsilon i \sim N(0, \delta^2)$

c. Sex ratio $_{ij} = \beta o + \beta_1$ (Time) + $\varepsilon_{ij} + u_j$ $\varepsilon ij \sim N(0, \delta^2)$; $u_j \sim N(0, \tau^2)$ for country j

d. Sex ratio $_{ij} = \beta o + \beta_1$ (Time) + $\varepsilon_{ij} + u_{1j} + u_{2j}$ Time $\varepsilon ij \sim N(0, \delta^2)$; $u_{1j} \sim N(0, \tau_1^2)$ for country j; $u_{2j} \sim N(0, \tau_2^2)$ for country j; Cov (u_{1i}, u_{2j}) = τ_{12}

† Unbalanced data (No. of countries different in different year) ‡ Balanced data (No. of countries same in each year) AIC gives an indication of which model to choose (lower is better)

A 6-i: Cancer of pancreas: regression models in three scenarios (A, B & C) with varying number of registry-populations (N): fixed-effects models (a and b); mixed effects model with random intercepts (c); and mixed effects model with random intercepts and slopes (d).

Models	Errors & AIC	βo	β_1	*SE* β_1	(95% CI)	intra- regional variance δ^2 (ε_{ij})	inter-regional variance τ^2 or τ_1^2	Variation in slopes τ_2^2	Cov (u_{1i}, u_{2j}) τ_{12}
(A) 1953-2007 (N= 3 to 113)†									
a. Fixed-Effects	ε 1106	1.86	-0.01	0.001	(-0.010, -0.006)				
b. Fixed Effects	ε 915	1.72	-0.01	0.001	(-0.013, -0.009)				
c. Random Intercepts	AIC 5360	1.98	-0.01	0.001	(-0.012, -0.008)	0.30*	τ^2 0.05*		
d. Random Intercepts & Slopes	AIC 5319	2.07	-0.01	0.002	(-0.016, -0.009)	0.29*	τ_1^2 0.45*	τ_2^2 0.0002*	τ_{12} -0.008*
(B) 1983-2007 (N = 76 to 113) †									
a. Fixed-Effects	ε 823	1.99	-0.01	0.002	(-0.014, -0.008)				
b. Fixed Effects	ε 665	1.77	-0.01	0.002	(-0.015, -0.009)				
c. Random Intercepts	AIC 4118	2.01	-0.01	0.002	(-0.015, -0.009)	0.27*	τ^2 0.05*		
d. Random Intercepts & Slopes	AIC 4098	2.06	-0.01	0.002	(-0.015, -0.009)	0.27*	τ_1^2 0.40*	τ_2^2 0.0001*	τ_{12} -0.006*
(C) 1983-2007 (N = 76) ‡									
a. Fixed-Effects	ε 498	2.07	-0.013	0.002	(-0.016, -0.010)				
b. Fixed Effects	ε 384	1.79	-0.013	0.002	(-0.016, -0.010)				
c. Random Intercepts	AIC 2572	2.07	-0.013	0.002	(-0.016, -0.010)	0.22*	τ^2 0.05*		
d. Random Intercepts & Slopes	AIC 2536	2.08	-0.013	0.002	(-0.017, -0.009)	0.21*	τ_1^2 0.53*	τ_2^2 0.0002*	τ_{12} -0.009*

a. Sex ratio $_i = \beta o + \beta_1$ (Time) + ε_i $\quad \varepsilon i \sim N(0, \delta^2)$ $\quad$ * p value < 0.005

b. Sex ratio $_i = \beta o + \beta_1$ (Time) + β_2 (country1) + β_3 (country2) ….. + β_n (country n) + ε_i $\quad \varepsilon i \sim N(0, \delta^2)$

c. Sex ratio $_{ij} = \beta o + \beta_1$ (Time) + $\varepsilon_{ij} + u_j$ $\quad \varepsilon ij \sim N(0, \delta^2)$; $\quad u_j \sim N(0, \tau^2)$ for country j

d. Sex ratio $_{ij} = \beta o + \beta_1$ (Time) + $\varepsilon_{ij} + u_{1j} + u_{2j}$ Time $\quad \varepsilon ij \sim N(0, \delta^2)$; $\quad u_{1j} \sim N(0, \tau_1^2)$ for country j; $\quad u_{2j} \sim N(0, \tau_2^2)$ for country j; $\quad$ Cov (u_{1i}, u_{2j}) = τ_{12}

† Unbalanced data (No. of countries different in different year) ‡ Balanced data (No. of countries same in each year) AIC gives an indication of which model to choose (lower is better)

A 6-j: Cancer of bone: regression models in three scenarios (A, B & C) with varying number of registry-populations (N): fixed-effects models (a and b); mixed effects model with random intercepts (c); and mixed effects model with random intercepts and slopes (d).

Models	Errors & AIC	βo	β_1	*SE* β_1	(95% CI)	intra- regional variance δ^2 (ε_{ij})	inter-regional variance τ^2 or τ_1^2	Variation in slopes τ_2^2	Cov (u_{1i}, u_{2j}) τ_{12}
(A) 1953-2007 (N= 3 to 113)†									
a. Fixed-Effects	ε 8129	1.42	0.006	0.003	(0.001, 0.012)				
b. Fixed Effects	ε 7550	1.31	0.001	0.003	(-0.006, 0.007)				
c. Random Intercepts	AIC 11271	1.51	0.004	0.003	(-0.002, 0.010)	2.72*	τ^2 0.075*		
d. Random Intercepts & Slopes	AIC 11274	1.50	0.004	0.003	(-0.001, 0.010)	2.71*	τ_1^2 0.067*	τ_2^2 ---	τ_{12} ---
(B) 1983-2007 (N = 76 to 113) †									
a. Fixed-Effects	ε 7633	1.83	-0.003	0.005	(-0.013, 0.008)				
b. Fixed Effects	ε 7063	1.58	-0.008	0.005	(-0.019, 0.003)				
c. Random Intercepts	AIC 9446	1.88	-0.004	0.005	(-0.014, 0.007)	3.18*	τ^2 0.064*		
d. Random Intercepts & Slopes	AIC 9817	2.02	-0.007	0.010	(-0.027, 0.014)	3.13*	τ_1^2 0.001	τ_2^2 0.009	τ_{12} -0.0032
(C) 1983-2007 (N = 76) ‡									
a. Fixed-Effects	ε 4567	1.85	-0.005	0.005	(-0.015, 0.006)				
b. Fixed Effects	ε 4339	1.46	-0.005	0.005	(-0.016, 0.005)				
c. Random Intercepts	AIC 6605	1.86	-0.005	0.005	(-0.015, 0.006)	2.62*	τ^2 0.015		
d. Random Intercepts & Slopes	AIC 6607	1.85	-0.005	0.006	(-0.017, 0.007)	2.59*	τ_1^2 0.624	τ_2^2 0.0004	τ_{12} -0.016

a. Sex ratio $_i = \beta o + \beta_1$ (Time) + ε_i $\varepsilon i \sim N(0, \delta^2)$ * p value < 0.005

b. Sex ratio $_i = \beta o + \beta_1$ (Time) + β_2 (country1) + β_3 (country2) ….. + β_n (country n) + ε_i $\varepsilon i \sim N(0, \delta^2)$

c. Sex ratio $_{ij} = \beta o + \beta_1$ (Time) + $\varepsilon_{ij} + u_j$ $\varepsilon ij \sim N(0, \delta^2)$; $u_j \sim N(0, \tau^2)$ for country j

d. Sex ratio $_{ij} = \beta o + \beta_1$ (Time) + $\varepsilon_{ij} + u_{1j} + u_{2j}$ Time $\varepsilon ij \sim N(0, \delta^2)$; $u_{1j} \sim N(0, \tau_1^2)$ for country j; $u_{2j} \sim N(0, \tau_2^2)$ for country j; Cov (u_{1i}, u_{2j}) = τ_{12}

† Unbalanced data (No. of countries different in different year) ‡ Balanced data (No. of countries same in each year) AIC gives an indication of which model to choose (lower is better)

A 6-k: Cancer of kidney: regression models in three scenarios (A, B & C) with varying number of registry-populations (N): fixed-effects models (a and b); mixed effects model with random intercepts (c); and mixed effects model with random intercepts and slopes (d).

Models	Errors & AIC	βo	β_1	*SE* β_1	(95% CI)	intra- regional variance δ^2 (ε_{ij})	inter-regional variance τ^2 or τ_1^2	Variation in slopes τ_2^2	Cov (u_{1i}, u_{2j}) τ_{12}
(A) 1953-2007 (N= 3 to 113)†									
a. Fixed-Effects	ε 3760	1.85	0.007	0.002	(0.004, 0.011)				
b. Fixed Effects	ε 3288	1.99	0.002	0.002	(-0.002, 0.006)				
c. Random Intercepts	AIC 9329	2.03	0.003	0.002	(-0.002, 0.007)	1.08*	τ^2 0.12*		
d. Random Intercepts & Slopes	AIC 9334	2.02	0.004	0.002	(-0.008, 0.007)	1.09*	τ_1^2 0.17*	τ_2^2 0.0001*	τ_{12} -0.003*
(B) 1983-2007 (N = 76 to 113) †									
a. Fixed-Effects	ε 3263	1.88	0.006	0.003	(-0.008, 0.013)				
b. Fixed Effects	ε 2834	1.84	0.004	0.003	(-0.002, 0.010)				
c. Random Intercepts	AIC 7749	1.95	0.005	0.003	(-0.001, 0.011)	1.17*	τ^2 0.12*		
d. Random Intercepts & Slopes	AIC 7820	1.89	0.006	0.003	(-0.0001, 0.122)	1.17*	τ_1^2 ----	τ_2^2 ----	τ_{12} ----
(C) 1983-2007 (N = 76) ‡									
a. Fixed-Effects	ε 1534	1.89	0.005	0.003	(-0.001, 0.010)				
b. Fixed Effects	ε 1351	1.82	0.005	0.003	(-0.001, 0.010)				
c. Random Intercepts	AIC 4827	1.89	0.004	0.003	(-0.001, 0.0101)	0.76*	τ^2 0.07*		
d. Random Intercepts & Slopes	AIC 4825	1.89	0.005	0.003	(-0.002, 0.0112)	0.75*	τ_1^2 0.39	τ_2^2 0.0002	τ_{12} -0.009*

a. Sex ratio $_i = \beta o + \beta_1$ (Time) + ε_i $\varepsilon i \sim N(0, \delta^2)$ * p value < 0.005

b. Sex ratio $_i = \beta o + \beta_1$ (Time) + β_2 (country1) + β_3 (country2) ….. + β_n (country n) + ε_i $\varepsilon i \sim N(0, \delta^2)$

c. Sex ratio $_{ij} = \beta o + \beta_1$ (Time) + $\varepsilon_{ij} + u_j$ $\varepsilon ij \sim N(0, \delta^2)$; $u_j \sim N(0, \tau^2)$ for country j

d. Sex ratio $_{ij} = \beta o + \beta_1$ (Time) + $\varepsilon_{ij} + u_{1j} + u_{2j}$ Time $\varepsilon ij \sim N(0, \delta^2)$; $u_{1j} \sim N(0, \tau_1^2)$ for country j; $u_{2j} \sim N(0, \tau_2^2)$ for country j; Cov (u_{1i}, u_{2j}) = τ_{12}

† Unbalanced data (No. of countries different in different year) ‡ Balanced data (No. of countries same in each year) AIC gives an indication of which model to choose (lower is better)

A 6-l: Leukemia: regression models in three scenarios (A, B & C) with varying number of registry-populations (N): fixed-effects models (a and b); mixed effects model with random intercepts (c); and mixed effects model with random intercepts and slopes (d).

Models	Errors & AIC	βo	β_1	*SE* β_1	(95% CI)	intra- regional variance δ^2 (ε_{ij})	inter-regional variance τ^2 or τ_1^2	Variation in slopes τ_2^2	Cov (u_{1i}, u_{2j}) τ_{12}
(A) 1953-2007 (N= 3 to 113)†									
a. Fixed-Effects	ε 954	1.62	-0.001	0.001	(-0.002, 0.001)				
b. Fixed Effects	ε 863	1.73	-0.002	0.001	(-0.004,-0.003)				
c. Random Intercepts	AIC 5111	1.66	-0.002	0.001	(-0.003,0.0002)	0.28*	τ^2 0.02*		
d. Random Intercepts & Slopes	AIC 5112	1.68	-0.002	0.001	(-0.004,0.0001)	0.28*	τ_1^2 0.05*	τ_2^2 0.00001	τ_{12} -0.001
(B) 1983-2007 (N = 76 to 113) †									
a. Fixed-Effects	ε 811	1.85	-0.0106	0.002	(-0009,-0.003)				
b. Fixed Effects	ε 727	1.85	-0.0106	0.002	(-0009,-0.003)				
c. Random Intercepts	AIC 4269	1.85	-0.0106	0.001	(-0009,-0.003)	0.29*	τ^2 0.02*		
d. Random Intercepts & Slopes	AIC 4272	1.85	-0.0106	0.002	(-0009,-0.003)	0.29*	τ_1^2 0.98	τ_2^2 0.00001	τ_{12} -0.001
(C) 1983-2007 (N = 76) ‡									
a. Fixed-Effects	ε 487	1.85	-0.0106	0.002	(-0009,-0.003)				
b. Fixed Effects	ε 435	1.85	-0.0106	0.002	(-0009,-0.003)				
c. Random Intercepts	AIC 2733	1.85	-0.0106	0.001	(-0009,-0.003)	0.25*	τ^2 0.25*		
d. Random Intercepts & Slopes	AIC 2729	1.85	-0.0106	0.001	(-0009,-0.002)	0.24*	τ_1^2 0.12*	τ_2^2 0.00002	τ_{12} -0.002

a. Sex ratio $_i = \beta o + \beta_1$ (Time) + ε_i $\varepsilon i \sim N(0, \delta^2)$ * p value < 0.005

b. Sex ratio $_i = \beta o + \beta_1$ (Time) + β_2 (country1) + β_3 (country2) ..... + β_n (country n) + ε_i $\varepsilon i \sim N(0, \delta^2)$

c. Sex ratio $_{ij} = \beta o + \beta_1$ (Time) + $\varepsilon_{ij} + u_j$ $\varepsilon ij \sim N(0, \delta^2)$; $u_j \sim N(0, \tau^2)$ for country j

d. Sex ratio $_{ij} = \beta o + \beta_1$ (Time) + $\varepsilon_{ij} + u_{1j} + u_{2j}$ Time $\varepsilon ij \sim N(0, \delta^2)$; $u_{1j} \sim N(0, \tau_1^2)$ for country j; $u_{2j} \sim N(0, \tau_2^2)$ for country j; Cov (u_{1i}, u_{2j}) = τ_{12}

† Unbalanced data (No. of countries different in different year) ‡ Balanced data (No. of countries same in each year) AIC gives an indication of which model to choose (lower is better)

A 6-m: Non-Hodgkin lymphoma: regression models in three scenarios (A, B & C) with varying number of registry-populations (N): fixed-effects models (a and b); mixed effects model with random intercepts (c); and mixed effects model with random intercepts and slopes (d).

Models	Errors & AIC	βo	β_1	*SE* β_1	(95% CI)	intra- regional variance δ^2 (ε_{ij})	inter-regional variance τ^2 or τ_1^2	Variation in slopes τ_2^2	Cov (u_{1i}, u_{2j}) τ_{12}
(A) 1953-2007 (N= 3 to 113)†									
a. Fixed-Effects	ε 1317	1.72	-0.004	0.001	(-0.006,-0.002)				
b. Fixed Effects	ε 1117	1.70	-0.006	0.001	(-0.008,-0.004)				
c. Random Intercepts	AIC 6004	1.79	-0.005	0.001	(-0.007,-0.003)	0.37*	τ^2 0.06*		
d. Random Intercepts & Slopes	AIC 6008	1.79	-0.006	0.001	(-0.008,-0.003)	0.37*	τ_1^2 0.07*	τ_2^2 ---	τ_{12} ----
(B) 1983-2007 (N = 76 to 113) †									
a. Fixed-Effects	ε 1134	1.76	-0.005	0.002	(-0.008,-0.001)				
b. Fixed Effects	ε 950	1.81	-0.007	0.002	(-0.011,-0.003)				
c. Random Intercepts	AIC 5029	1.83	-0.006	0.002	(-0.009,-0.003)	0.39*	τ^2 0.06*		
d. Random Intercepts & Slopes	AIC 5378	1.83	-0.006	0.005	(-0.015,0.003)	0.39*	τ_1^2 ---	τ_2^2 ---	τ_{12} ---
(C) 1983-2007 (N = 76) ‡									
a. Fixed-Effects	ε 369	1.76	-0.006	0.001	(-0.009,-0.003)				
b. Fixed Effects	ε 315	1.76	-0.006	0.001	(-0.008,-0.003)				
c. Random Intercepts	AIC 2162	1.76	-0.006	0.001	(-0.008,-0.003)	0.18*	τ^2 0.02*		
d. Random Intercepts & Slopes	AIC 2144	1.76	-0.006	0.001	(-0.009,-0.002)	0.17*	τ_1^2 0.21	τ_2^2 0.0001	τ_{12} -0.003*

a. Sex ratio $_i = \beta o + \beta_1$ (Time) + ε_i $\quad \varepsilon i \sim N(0, \delta^2)$ $\quad$ * p value < 0.005

b. Sex ratio $_i = \beta o + \beta_1$ (Time) + β_2 (country1) + β_3 (country2) ….. + β_n (country n) + ε_i $\quad \varepsilon i \sim N(0, \delta^2)$

c. Sex ratio $_{ij} = \beta o + \beta_1$ (Time) + $\varepsilon_{ij} + u_j$ $\quad \varepsilon ij \sim N(0, \delta^2)$; $\quad u_j \sim N(0, \tau^2)$ for country j

d. Sex ratio $_{ij} = \beta o + \beta_1$ (Time) + $\varepsilon_{ij} + u_{1j} + u_{2j}$ Time $\quad \varepsilon ij \sim N(0, \delta^2)$; $\quad u_{1j} \sim N(0, \tau_1^2)$ for country j; $\quad u_{2j} \sim N(0, \tau_2^2)$ for country j; $\quad$ Cov (u_{1i}, u_{2j}) = τ_{12}

† Unbalanced data (No. of countries different in different year) ‡ Balanced data (No. of countries same in each year) AIC gives an indication of which model to choose (lower is better)

A 6-n: Multiple myeloma: regression models in three scenarios (A, B & C) with varying number of registry-populations (N): fixed-effects models (a and b); mixed effects model with random intercepts (c); and mixed effects model with random intercepts and slopes (d).

Models	Errors & AIC	βo	β_1	*SE* β_1	(95% CI)	intra- regional variance δ^2 (ε_{ij})	inter-regional variance τ^2 or τ_1^2	Variation in slopes τ_2^2	Cov (u_{1i}, u_{2j}) τ_{12}
(A) 1953-2007 (N= 3 to 113)†									
a. Fixed-Effects	ε 2312	1.48	0.002	0.001	(-0.001,0.005)				
b. Fixed Effects	ε 2127		0.001	0.002	(-0.002,0.004)				
c. Random Intercepts	AIC 7867	1.49	0.001	0.002	(-0.002,0.004)	0.71*	τ^2 0.04*		
d. Random Intercepts & Slopes	AIC 7872	1.49	0.002	0.002	(-0.002,0.005)	0.71*	τ_1^2 0.13*	τ_2^2 0.00003	τ_{12} -0.002
(B) 1983-2007 (N = 76 to 113) †									
a. Fixed-Effects	ε 1845	1.48	0.002	0.002	(-0.003,0.006)				
b. Fixed Effects	ε 1673	1.61	0.001	0.002	(-0.004,0.005)				
c. Random Intercepts	AIC 6336	1.51	0.001	0.002	(-0.004,0.006)	0.69*	τ^2 0.04*		
d. Random Intercepts & Slopes	AIC	1.53	0.001	0.003	(-0.005,0.007)	0.69*	τ_1^2 0.001*	τ_2^2 0.002	τ_{12} 0.001
(C) 1983-2007 (N = 76) ‡									
a. Fixed-Effects	ε 856	1.51	0.001	0.002	(-0.004,0.005)				
b. Fixed Effects	ε 796	1.61	0.001	0.002	(-0.004,0.005)				
c. Random Intercepts	AIC	1.52	0.001	0.002	(-0.004,0.005)	0.45*	τ^2 0.014		
d. Random Intercepts & Slopes	AIC	1.52	0.004	0.002	(-0.004,0.005)	0.45*	τ_1^2 0.23	τ_2^2 0.0001	τ_{12} -0.004

a. Sex ratio $_i = \beta o + \beta_1 (\text{Time}) + \varepsilon_i$ $\varepsilon i \sim N(0, \delta^2)$ * p value < 0.005

b. Sex ratio $_i = \beta o + \beta_1 (\text{Time}) + \beta_2 (\text{country1}) + \beta_3 (\text{country2}) \ldots\ldots + \beta_n (\text{country n}) + \varepsilon_i$ $\varepsilon i \sim N(0, \delta^2)$

c. Sex ratio $_{ij} = \beta o + \beta_1 (\text{Time}) + \varepsilon_{ij} + u_j$ $\varepsilon ij \sim N(0, \delta^2)$; $u_j \sim N(0, \tau^2)$ for country j

d. Sex ratio $_{ij} = \beta o + \beta_1 (\text{Time}) + \varepsilon_{ij} + u_{1j} + u_{2j}$ Time $\varepsilon ij \sim N(0, \delta^2)$; $u_{1j} \sim N(0, \tau_1^2)$ for country j; $u_{2j} \sim N(0, \tau_2^2)$ for country j; Cov (u_{1i}, u_{2j}) = τ_{12}

† Unbalanced data (No. of countries different in different year) ‡ Balanced data (No. of countries same in each year) AIC gives an indication of which model to choose (lower is better)

A 6-o: Cancer of rectum and anus : regression models in three scenarios (A, B & C) with varying number of registry-populations (N): fixed-effects models (a and b); mixed effects model with random intercepts (c); and mixed effects model with random intercepts and slopes (d).

Models	Errors & AIC	βo	β_1	*SE* β_1	(95% CI)	intra- regional variance δ^2 (ε_{ij})	inter-regional variance τ^2 or τ_1^2	Variation in slopes τ_2^2	Cov (u_{1i}, u_{2j}) τ_{12}
(A) 1953-2007 (N= 3 to 113)†									
a. Fixed-Effects	ε 672	1.43	0.006	0.001	(0.004, 0.007)				
b. Fixed Effects	ε 505	1.32	0.005	0.001	(0.003, 0.006)				
c. Random Intercepts	AIC 3550	1.47	0.005	0.001	(0.004, 0.007)	0.16*	τ^2 0.05*		
d. Random Intercepts & Slopes	AIC 3519	1.47	0.005	0.001	(0.003, 0.006)	0.16*	τ_1^2 0.11*	τ_2^2 0.00002*	τ_{12} -0.002*
(B) 1983-2007 (N = 76 to 113) †									
a. Fixed-Effects	ε 579	1.52	0.004	0.001	(0.001, 0.006)				
b. Fixed Effects	ε 418	1.35	0.004	0.001	(0.002, 0.007)				
c. Random Intercepts	AIC 2987	1.51	0.004	0.001	(0.002, 0.007)	0.17*	τ^2 0.06*		
d. Random Intercepts & Slopes	AIC 2988	1.51	0.004	0.001	(0.001, 0.007)	0.17*	τ_1^2 0.01*	τ_2^2 0.00003*	τ_{12} -0.001*
(C) 1983-2007 (N = 76) ‡									
a. Fixed-Effects	ε 359	1.53	0.004	0.001	(0.001, 0.007)				
b. Fixed Effects	ε 253	1.36	0.004	0.001	(0.002, 0.006)				
c. Random Intercepts	AIC 1823	1.53	0.004	0.001	(0.001, 0.006)	0.14*	τ^2 0.05*		
d. Random Intercepts & Slopes	AIC 1817	1.53	0.004	0.002	(0.001, 0.007)	0.14*	τ_1^2 0.16*	τ_2^2 0.001*	τ_{12} -0.003*

a. Sex ratio $_i = \beta o + \beta_1$ (Time) + ε_i $\quad \varepsilon i \sim N(0, \delta^2)$ $\quad$ * p value < 0.005

b. Sex ratio $_i = \beta o + \beta_1$ (Time) + β_2 (country1) + β_3 (country2) ….. + β_n (country n) + ε_i $\quad \varepsilon i \sim N(0, \delta^2)$

c. Sex ratio $_{ij} = \beta o + \beta_1$ (Time) + $\varepsilon_{ij} + u_j$ $\quad \varepsilon ij \sim N(0, \delta^2)$; $\quad u_j \sim N(0, \tau^2)$ for country j

d. Sex ratio $_{ij} = \beta o + \beta_1$ (Time) + $\varepsilon_{ij} + u_{1j} + u_{2j}$ Time $\quad \varepsilon ij \sim N(0, \delta^2)$; $\quad u_{1j} \sim N(0, \tau_1^2)$ for country j; $\quad u_{2j} \sim N(0, \tau_2^2)$ for country j; $\quad$ Cov (u_{1i}, u_{2j}) = τ_{12}

† Unbalanced data (No. of countries different in different year) ‡ Balanced data (No. of countries same in each year) AIC gives an indication of which model to choose (lower is better)

A 6-p: Cancer of brain: regression models in three scenarios (A, B & C) with varying number of registry-populations (N): fixed-effects models (a and b); mixed effects model with random intercepts (c); and mixed effects model with random intercepts and slopes (d).

Models	Errors & AIC	βo	β_1	*SE* β_1	(95% CI)	intra- regional variance δ^2 (ε_{ij})	inter-regional variance τ^2 or τ_1^2	Variation in slopes τ_2^2	Cov (u_{1i}, u_{2j}) τ_{12}
(A) 1953-2007 (N= 3 to 113)†									
a. Fixed-Effects	ε 3202	1.43	0.002	0.002	(-0.001,0.005)				
b. Fixed Effects	ε 2940	1.50	-0.001	0.002	(-0.001,0.003)				
c. Random Intercepts	AIC 8932	1.51	0.0001	0.0018	(0.956,-0.003)	0.97*	τ^2 0.52*		
d. Random Intercepts & Slopes	AIC 8678	1.34	0.0040	0.0055	(-0.007,0.0149)	0.85*	τ_1^2 4.44*	τ_2^2 0.003*	τ_{12} -0.115*
(B) 1983-2007 (N = 76 to 113) †									
a. Fixed-Effects	ε 2961	1.41	0.002	0.003	(-0.004,0.008)				
b. Fixed Effects	ε 2713	1.39	0.002	0.003	(-0.004,0.008)				
c. Random Intercepts	AIC 7584	1.43	0.002	0.003	(-0.004,0.008)	1.12*	τ^2 0.05*		
d. Random Intercepts & Slopes	AIC 7355	1.39	0.003	0.006	(-0.009,0.014)	0.97*	τ_1^2 4.83*	τ_2^2 0.003*	τ_{12} -0.124*
(C) 1983-2007 (N = 76) ‡									
a. Fixed-Effects	ε 980	1.65	-0.004	0.002	(-0.008,0.001)				
b. Fixed Effects	ε 905	1.62	-0.004	0.002	(-0.008,0.001)				
c. Random Intercepts	AIC 4055	1.65	-0.004	0.002	(-0.008,0.0001)	0.51*	τ^2 0.51*		
d. Random Intercepts & Slopes	AIC 4055	1.65	-0.004	0.002	(-0.008,0.0011)	0.51*	τ_1^2 0.17*	τ_2^2 0.00003*	τ_{12} -0.002*

a. Sex ratio $_i = \beta o + \beta_1 (\text{Time}) + \varepsilon_i$ $\varepsilon i \sim N(0, \delta^2)$ * p value < 0.005

b. Sex ratio $_i = \beta o + \beta_1 (\text{Time}) + \beta_2 (\text{country1}) + \beta_3 (\text{country2}) \ldots\ldots + \beta_n (\text{country n}) + \varepsilon_i$ $\varepsilon i \sim N(0, \delta^2)$

c. Sex ratio $_{ij} = \beta o + \beta_1 (\text{Time}) + \varepsilon_{ij} + u_j$ $\varepsilon ij \sim N(0, \delta^2)$; $u_j \sim N(0, \tau^2)$ for country j

d. Sex ratio $_{ij} = \beta o + \beta_1 (\text{Time}) + \varepsilon_{ij} + u_{1j} + u_{2j}$ Time $\varepsilon ij \sim N(0, \delta^2)$; $u_{1j} \sim N(0, \tau_1^2)$ for country j; $u_{2j} \sim N(0, \tau_2^2)$ for country j; Cov (u_{1i}, u_{2j}) = τ_{12}

† Unbalanced data (No. of countries different in different year) ‡ Balanced data (No. of countries same in each year) AIC gives an indication of which model to choose (lower is better)

A 6-q: Cancer of Eye: regression models in three scenarios (A, B & C) with varying number of registry-populations (N): fixed-effects models (a and b); mixed effects model with random intercepts (c); and mixed effects model with random intercepts and slopes (d).

Models	Errors & AIC	βo	β_1	*SE* β_1	(95% CI)	intra- regional variance δ^2 (ε_{ij})	inter-regional variance τ^2 or τ_1^2	Variation in slopes τ_2^2	Cov (u_{1i}, u_{2j}) τ_{12}
(A) 1953-2007 (N= 3 to 113)†									
a. Fixed-Effects	ε 6592	1.23	0.008	0.003	(0.001,0.012)				
b. Fixed Effects	ε 5993	1.09	0.007	0.003	(0.0001,0.012)				
c. Random Intercepts	AIC 9934	1.31	0.008	0.003	(-0.0001,0.0114)	2.36*	τ^2 0.13*		
d. Random Intercepts & Slopes	AIC 9921	1.26	0.008	0.004	(-0.0004,0.0137)	2.32*	τ_1^2 0.29	τ_2^2 0.0004*	τ_{12} -0.010
(B) 1983-2007 (N = 76 to 113) †									
a. Fixed-Effects	ε 5816	1.23	0.008	0.005	(-0.002, 0.018)				
b. Fixed Effects	ε 5218	0.89	0.011	0.005	(-0.001, 0.021)				
c. Random Intercepts	AIC 8204	1.14	0.009	0.005	(-0.0005, 0.0193)	2.58*	τ^2 0.14*		
d. Random Intercepts & Slopes	AIC 8231	1.18	0.008	0.005	(-0.0025, 0.0189)	2.55*	τ_1^2 0.29	τ_2^2 0.0007	τ_{12} -0.009
(C) 1983-2007 (N = 76) ‡									
a. Fixed-Effects	ε 3982	1.16	0.010	0.005	(0.000, 0.021)				
b. Fixed Effects	ε 3603	0.94	0.010	0.005	(0.002, 0.021)				
c. Random Intercepts	AIC 5986	1.15	0.010	0.005	(-0.0003, 0.2076)	2.37*	τ^2 0.12*		
d. Random Intercepts & Slopes	AIC 5978	1.20	0.009	0.006	(-0.0003, 0.0214)	2.33*	τ_1^2 0.57	τ_2^2 0.0007	τ_{12} -0203

a. Sex ratio $_i = \beta o + \beta_1$ (Time) + ε_i $\quad \varepsilon i \sim N(0, \delta^2)$ $\quad$ * p value < 0.005

b. Sex ratio $_i = \beta o + \beta_1$ (Time) + β_2 (country1) + β_3 (country2) ….. + β_n (country n) + ε_i $\quad \varepsilon i \sim N(0, \delta^2)$

c. Sex ratio $_{ij} = \beta o + \beta_1$ (Time) + $\varepsilon_{ij} + u_j$ $\quad \varepsilon ij \sim N(0, \delta^2)$; $\quad u_j \sim N(0, \tau^2)$ for country j

d. Sex ratio $_{ij} = \beta o + \beta_1$ (Time) + $\varepsilon_{ij} + u_{1j} + u_{2j}$ Time $\quad \varepsilon ij \sim N(0, \delta^2)$; $\quad u_{1j} \sim N(0, \tau_1^2)$ for country j; $\quad u_{2j} \sim N(0, \tau_2^2)$ for country j; $\quad$ Cov (u_{1i}, u_{2j}) = τ_{12}

† Unbalanced data (No. of countries different in different year) ‡ Balanced data (No. of countries same in each year) AIC gives an indication of which model to choose (lower is better)

A 6-r: Cancer of colon : regression models in three scenarios (A, B & C) with varying number of registry-populations (N): fixed-effects models (a and b); mixed effects model with random intercepts (c); and mixed effects model with random intercepts and slopes (d).

Models	Errors & AIC	βo	β_1	*SE* β_1	(95% CI)	intra- regional variance δ^2 (ε_{ij})	inter-regional variance τ^2 or τ_1^2	Variation in slopes τ_2^2	Cov (u_{1i}, u_{2j}) τ_{12}
(A) 1953-2007 (N= 3 to 113)†									
a. Fixed-Effects	ε 340	1.00	0.008	0.001	(0.007, 0.009)				
b. Fixed Effects	ε 256	1.00	0.007	0.001	(0.006, 0.008)				
c. Random Intercepts	AIC 1399	1.05	0.007	0.005	(0.005, 0.008)	0.084*	τ^2 0.03*		
d. Random Intercepts & Slopes	AIC 1342	1.06	0.007	0.001	(0.004, 0.009)	0.079*	τ_1^2 0.16*	τ_2^2 0.0001	τ_{12} -0.003
(B) 1983-2007 (N = 76 to 113) †									
a. Fixed-Effects	ε 297	1.16	0.004	0.001	(0.002, 0.006)				
b. Fixed Effects	ε 216	1.10	0.004	0.001	(0.002, 0.006)				
c. Random Intercepts	AIC 1305	1.17	0.041	0.009	(0.002, 0.006)	0.088*	τ^2 0.03*		
d. Random Intercepts & Slopes	AIC 1231	1.15	0.005	0.001	(0.002, 0.007)	0.082*	τ_1^2 0.29*	τ_2^2 0.0001	τ_{12} -0.006*
(C) 1983-2007 (N = 76) ‡									
a. Fixed-Effects	ε 202	1.21	0.003	0.001	(0.001, 0.005)				
b. Fixed Effects	ε 151	1.14	0.003	0.001	(0.001, 0.005)				
c. Random Intercepts	AIC 853	1.21	0.003	0.001	(0.001, 0.005)	0.09*	τ^2 0.02*		
d. Random Intercepts & Slopes	AIC 773	1.21	0.003	0.002	(-0.0002, 0.006)	0.08*	τ_1^2 0.31*	τ_2^2 0.0001*	τ_{12} -0.007*

a. Sex ratio $_i = \beta o + \beta_1 (\text{Time}) + \varepsilon_i$ $\varepsilon i \sim N(0, \delta^2)$ * p value < 0.005

b. Sex ratio $_i = \beta o + \beta_1 (\text{Time}) + \beta_2 (\text{country1}) + \beta_3 (\text{country2}) \ldots\ldots + \beta_n (\text{country n}) + \varepsilon_i$ $\varepsilon i \sim N(0, \delta^2)$

c. Sex ratio $_{ij} = \beta o + \beta_1 (\text{Time}) + \varepsilon_{ij} + u_j$ $\varepsilon ij \sim N(0, \delta^2)$; $u_j \sim N(0, \tau^2)$ for country j

d. Sex ratio $_{ij} = \beta o + \beta_1 (\text{Time}) + \varepsilon_{ij} + u_{1j} + u_{2j}$ Time $\varepsilon ij \sim N(0, \delta^2)$; $u_{1j} \sim N(0, \tau_1^2)$ for country j; $u_{2j} \sim N(0, \tau_2^2)$ for country j; Cov (u_{1i}, u_{2j}) = τ_{12}

† Unbalanced data (No. of countries different in different year) ‡ Balanced data (No. of countries same in each year) AIC gives an indication of which model to choose (lower is better)

A 6-s: Cancer of skin (melanoma): regression models in three scenarios (A, B & C) with varying number of registry-populations (N): fixed-effects models (a and b); mixed effects model with random intercepts (c); and mixed effects model with random intercepts and slopes (d).

Models	Errors & AIC	βo	β_1	*SE* β_1	(95% CI)	intra- regional variance δ^2 (ε_{ij})	inter-regional variance τ^2 or τ_1^2	Variation in slopes τ_2^2	Cov (u_{1i}, u_{2j}) τ_{12}
(A) 1953-2007 (N= 3 to 113)†									
a. Fixed-Effects	ε 1914	0.87	0.005	0.001	(0.003, 0.008)				
b. Fixed Effects	ε 1571	0.97	0.005	0.001	(0.002, 0.007)				
c. Random Intercepts	AIC 6946	0.89	0.005	0.001	(0.002, 0.007)	0.54*	τ^2 0.09*		
d. Random Intercepts & Slopes	AIC 6917	0.94	0.004	0.002	(-0.001, 0.009)	0.52*	τ_1^2 1.05*	τ_2^2 0.0003*	τ_{12} -0.018*
(B) 1983-2007 (N = 76 to 113) †									
a. Fixed-Effects	ε 1712	0.95	0.004	0.002	(-0.001, 0.008)				
b. Fixed Effects	ε 1409	1.01	0.004	0.002	(-0.001, 0.008)				
c. Random Intercepts	AIC 5920	0.94	0.004	0.002	(-0.001, 0.008)	0.59*	τ^2 0.09*		
d. Random Intercepts & Slopes	AIC 5884	0.96	0.003	0.003	(-0.003, 0.009)	0.57*	τ_1^2 1.52*	τ_2^2 0.001*	τ_{12} -0.027*
(C) 1983-2007 (N = 76) ‡									
a. Fixed-Effects	ε 1187	0.91	0.005	0.003	(0.001, 0.011)				
b. Fixed Effects	ε 958	0.96	0.005	0.002	(0.001, 0.010)				
c. Random Intercepts	AIC 4189	0.92	0.005	0.002	(0.0003, 0.001)	0.55*	τ^2 0.11*		
d. Random Intercepts & Slopes	AIC 4166	0.92	0.005	0.003	(-0.002, 0.012)	0.53*	τ_1^2 1.14*	τ_2^2 0.0004*	τ_{12} -0.022*

a. Sex ratio $_i = \beta o + \beta_1$ (Time) + ε_i $\quad \varepsilon i \sim N(0, \delta^2)$ $\quad$ * p value < 0.005

b. Sex ratio $_i = \beta o + \beta_1$ (Time) + β_2 (country1) + β_3 (country2) ….. + β_n (country n) + ε_i $\quad \varepsilon i \sim N(0, \delta^2)$

c. Sex ratio $_{ij} = \beta o + \beta_1$ (Time) + $\varepsilon_{ij} + u_j$ $\quad \varepsilon ij \sim N(0, \delta^2)$; $\quad u_j \sim N(0, \tau^2)$ for country j

d. Sex ratio $_{ij} = \beta o + \beta_1$ (Time) + $\varepsilon_{ij} + u_{1j} + u_{2j}$ Time $\quad \varepsilon ij \sim N(0, \delta^2)$; $\quad u_{1j} \sim N(0, \tau_1^2)$ for country j; $\quad u_{2j} \sim N(0, \tau_2^2)$ for country j; $\quad$ Cov (u_{1i}, u_{2j}) = τ_{12}

† Unbalanced data (No. of countries different in different year) ‡ Balanced data (No. of countries same in each year) AIC gives an indication of which model to choose (lower is better)

A 6-t: Cancer of gallbladder : regression models in three scenarios (A, B & C) with varying number of registry-populations (N): fixed-effects models (a and b); mixed effects model with random intercepts (c); and mixed effects model with random intercepts and slopes (d).

Models	Errors & AIC	βo	β_1	*SE* β_1	(95% CI)	intra- regional variance δ^2 (ε_{ij})	inter-regional variance τ^2 or τ_1^2	Variation in slopes τ_2^2	Cov (u_{1i}, u_{2j}) τ_{12}
(A) 1953-2007 (N= 3 to 113)†									
a. Fixed-Effects	ε 1382	0.58	0.10	0.001	(0.008, 0.012)				
b. Fixed Effects	ε 1173	0.65	0.01	0.001	(0.006, 0.011)				
c. Random Intercepts	AIC 6107	0.62	0.01	0.001	(0.006, 0.011)	0.39*	τ^2 0.06*		
d. Random Intercepts & Slopes	AIC 6101	0.64	0.01	0.001	(0.005, 0.011)	0.38*	τ_1^2 0.17*	τ_2^2 0.0001*	τ_{12} -0.003*
(B) 1983-2007 (N = 76 to 113) †									
a. Fixed-Effects	ε 1229	0.61	0.01	0.002	(0.005, 0.013)				
b. Fixed Effects	ε 1043	0.65	0.01	0.002	(0.006, 0.013)				
c. Random Intercepts	AIC 5217	0.59	0.01	0.002	(0.006, 0.013)	0.43*	τ^2 0.05*		
d. Random Intercepts & Slopes	AIC 5220	0.59	0.01	0.002	0.001, 0.013)	0.43*	τ_1^2 0.09	τ_2^2 0.0001	τ_{12} -0.001
(C) 1983-2007 (N = 76) ‡									
a. Fixed-Effects	ε 788	0.64	0.01	0.002	(0.004, 0.013)				
b. Fixed Effects	ε 678	0.71	0.01	0.002	(0.004, 0.012)				
c. Random Intercepts	AIC 3566	0.64	0.01	0.002	(0.004, 0.012)	0.38*	τ^2 0.04*		
d. Random Intercepts & Slopes	AIC 3569	0.65	0.01	0.002	(0.004, 0.012)	0.38*	τ_1^2 0.16	τ_2^2 0.00004	τ_{12} -0.002

a. Sex ratio $_i = \beta o + \beta_1$ (Time) + ε_i $\varepsilon i \sim N(0, \delta^2)$ * p value < 0.005

b. Sex ratio $_i = \beta o + \beta_1$ (Time) + β_2 (country1) + β_3 (country2) ….. + β_n (country n) + ε_i $\varepsilon i \sim N(0, \delta^2)$

c. Sex ratio $_{ij} = \beta o + \beta_1$ (Time) + $\varepsilon_{ij} + u_j$ $\varepsilon ij \sim N(0, \delta^2)$; $u_j \sim N(0, \tau^2)$ for country j

d. Sex ratio $_{ij} = \beta o + \beta_1$ (Time) + $\varepsilon_{ij} + u_{1j} + u_{2j}$ Time $\varepsilon ij \sim N(0, \delta^2)$; $u_{1j} \sim N(0, \tau_1^2)$ for country j; $u_{2j} \sim N(0, \tau_2^2)$ for country j; Cov (u_{1i}, u_{2j}) = τ_{12}

† Unbalanced data (No. of countries different in different year) ‡ Balanced data (No. of countries same in each year) AIC gives an indication of which model to choose (lower is better)

A 6-u: Cancer of thyroid: regression models in three scenarios (A, B & C) with varying number of registry-populations (N): fixed-effects models (a and b); mixed effects model with random intercepts (c); and mixed effects model with random intercepts and slopes (d).

Models	Errors & AIC	βo	β_1	*SE* β_1	(95% CI)	intra- regional variance δ^2 (ε_{ij})	inter-regional variance τ^2 or τ_1^2	Variation in slopes τ_2^2	Cov (u_{1i}, u_{2j}) τ_{12}
(A) 1953-2007 (N= 3 to 113)†									
a. Fixed-Effects	ε 189	0.51	-0.003	0.0001	(-0.004,-0.002)				
b. Fixed Effects	ε 172	0.49	-0.003	0.0001	(-0.004,-0.003)				
c. Random Intercepts	AIC 17	0.52	-0.003	0.0004	(-0.004,-0.003)	0.06*	τ^2 0.004*		
d. Random Intercepts & Slopes	AIC 18	0.53	-0.003	0.0005	(-0.005,-0.003)	0.06*	τ_1^2 0.11*	τ_2^2 0.000003	τ_{12} -0.0002
(B) 1983-2007 (N = 76 to 113) †									
a. Fixed-Effects	ε 122	0.52	-0.003	0.0001	(-0.004,-0.002)				
b. Fixed Effects	ε 105	0.50	-0.004	0.0001	(-0.004,-0.003)				
c. Random Intercepts	AIC 622	0.53	-0.004	0.0001	(-0.004,-0.003)	0.04*	τ^2 0.005*		
d. Random Intercepts & Slopes	AIC 625	0.54	-0.004	0.0001	(-0.005,-0.002)	0.04*	τ_1^2 0.03*	τ_2^2 0.00001	τ_{12} -0005
(C) 1983-2007 (N = 76) ‡									
a. Fixed-Effects	ε 73	0.51	-0.004	0.0001	(-0.005,-0.003)				
b. Fixed Effects	ε 64	0.51	-0.004	0.0001	(-0.005,-0.003)				
c. Random Intercepts	AIC 881	0.54	-0.004	0.0001	(-0.005,-0.003)	0.04*	τ^2 0.004*		
d. Random Intercepts & Slopes	AIC 904	0.54	-0.004	0.0001	(-0.005,-0.003)	0.03*	τ_1^2 0.03*	τ_2^2 -0.001	τ_{12} -00001

a. Sex ratio $_i = \beta o + \beta_1$ (Time) + ε_i $\varepsilon i \sim N(0, \delta^2)$ * p value < 0.005

b. Sex ratio $_i = \beta o + \beta_1$ (Time) + β_2 (country1) + β_3 (country2) ….. + β_n (country n) + ε_i $\varepsilon i \sim N(0, \delta^2)$

c. Sex ratio $_{ij} = \beta o + \beta_1$ (Time) + $\varepsilon_{ij} + u_j$ $\varepsilon ij \sim N(0, \delta^2)$; $u_j \sim N(0, \tau^2)$ for country j

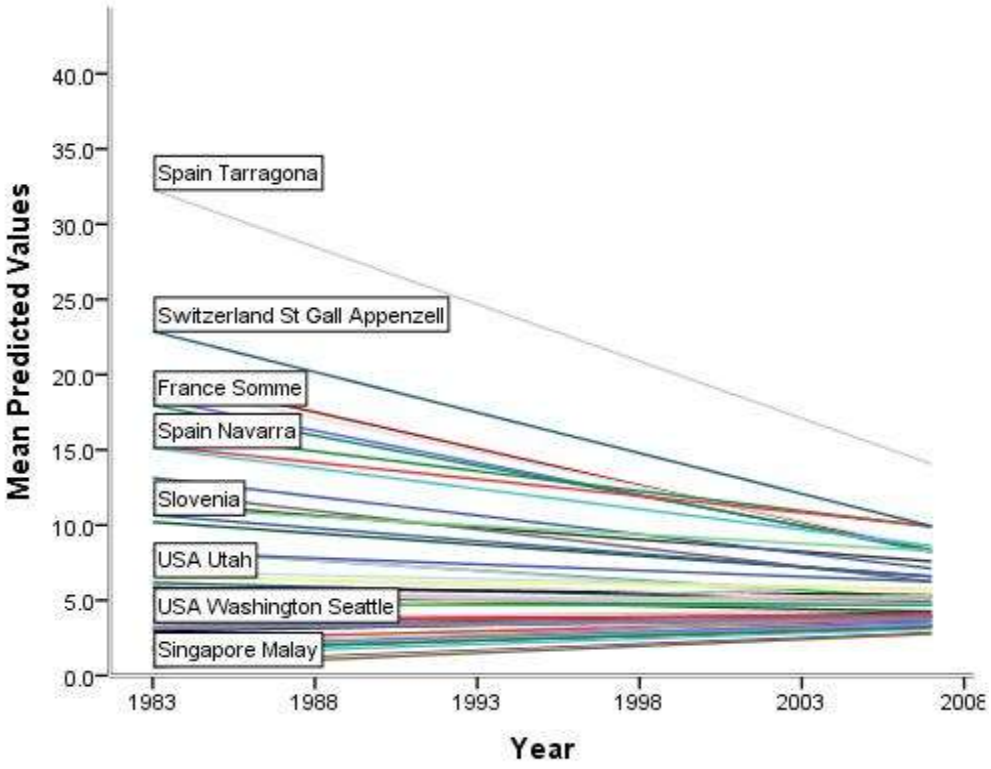
d. Sex ratio $_{ij} = \beta o + \beta_1$ (Time) + $\varepsilon_{ij} + u_{1j} + u_{2j}$ Time $\varepsilon ij \sim N(0, \delta^2)$; $u_{1j} \sim N(0, \tau_1^2)$ for country j; $u_{2j} \sim N(0, \tau_2^2)$ for country j; Cov (u_{1i}, u_{2j}) = τ_{12}

† Unbalanced data (No. of countries different in different year) ‡ Balanced data (No. of countries same in each year) AIC gives an indication of which model to choose (lower is better)

Appendix 7: Plots of worldwide SR magnitude (mean predicted values) form 1983-2007 for 11 types of cancers (A 7-a to A 7-h)

- Basic notations are used in the explanation of the graphs e.g., δ^2 for intra-regional variations and τ_1^2 for between regional variations.
- Formal notations for intra and inter-regional variations are explained in appendix 2.
- Observations on the trajectories over time are explained below the figures (A 7-a to A7-h)
- Intra- and inter- regional variations, variation in trajectories (slopes) and covariance of slopes in appendix 6 can be visually absorbed from the graphs in appendix 7.

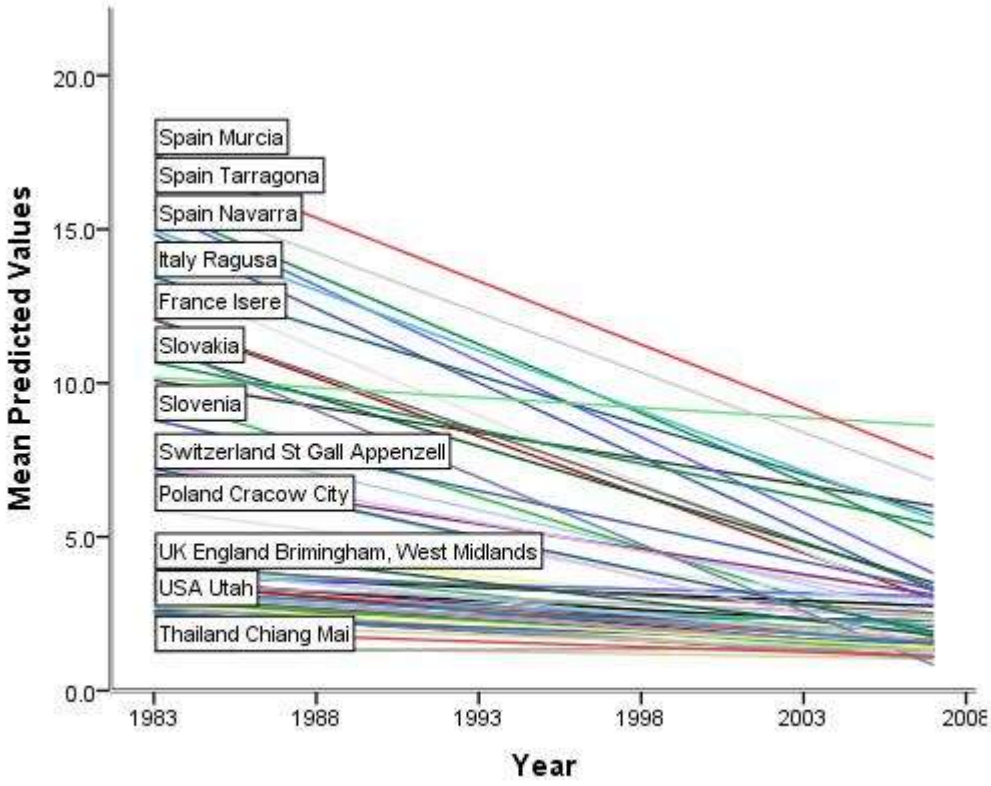
A 7-a Esophageal cancer: mean predicted value of sex ratios plotted against time 1983-2007



RI & RS model based on 76 cancer registries δ^2 (22.24) < τ_1^2 (132.53).

Cancer of esophagus showed similar pattern as cancer of larynx (page--) i.e., the baseline SR on average was among the highest i.e., 7.83 with average annual decrease of 0.052 (Table 8-14). After cancer of larynx, within- and between- variation of the registries was highest (within δ^2= 22.24, and between τ_1^2 = 132.53). The steepest declines in 25 years was seen in Tarragona (Spain), St. Gall Appenzell (Switzerland), Doubs and Somme in France. The lowest SR on average was again observed in Singapore, however the trajectory showed an increasing trend. The covariance between the trajectories was also significant (Appendix 6: A 6-c). The covariance of trajectories is also evident from Figure, where the registries with highest baseline SR have steepest declines in 25 years.

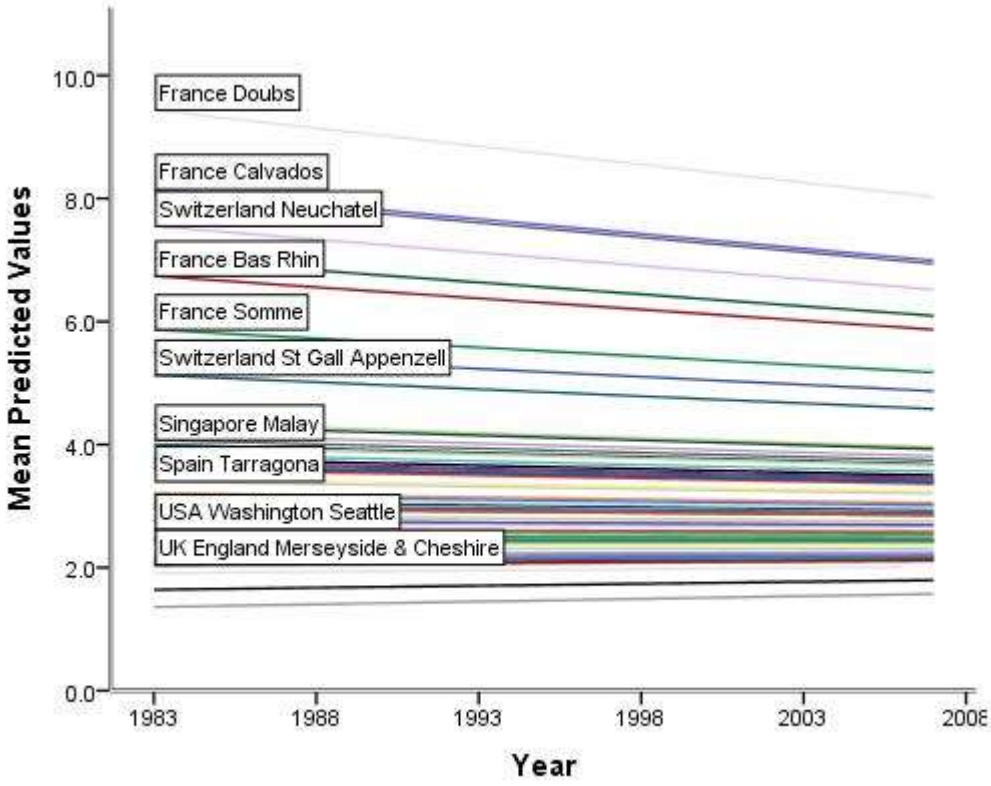
A 7-b Lung cancer: mean predicted value of sex ratios plotted against time 1983-2007



RI & RS model based on 76 cancer registries where δ^2 (1.57) < τ_1^2 (69.42).

After cancer of larynx, lung cancer showed the highest baseline SR of almost 10 in RI & RS model in 1983. Table 8-14 shows that the average annual rate of decrease of lung cancer from 1983 to 2007 was 0.141 and this was a significant decrease (-0.174, -0.109). The figure above shows some of the registries with their baseline SR on average, and variations in their trajectories. It is interesting to observe from both table 7-14 and figure above, that compared to cancers of larynx and esophagus, lung cancers shows a very little within-registry variance (within- variance: δ^2 (ε_{ij}) = 1.57, compared to between- variance: τ_1^2 = 69.42). For both cancers of larynx and esophagus, the within- component of trajectories was very high while it was still much lower than between- variations of SR (Table 7-14). Significant covariance of trajectories was also observed [Cov (u_{1i} , u_{2j}) τ_{12} = -1.13].

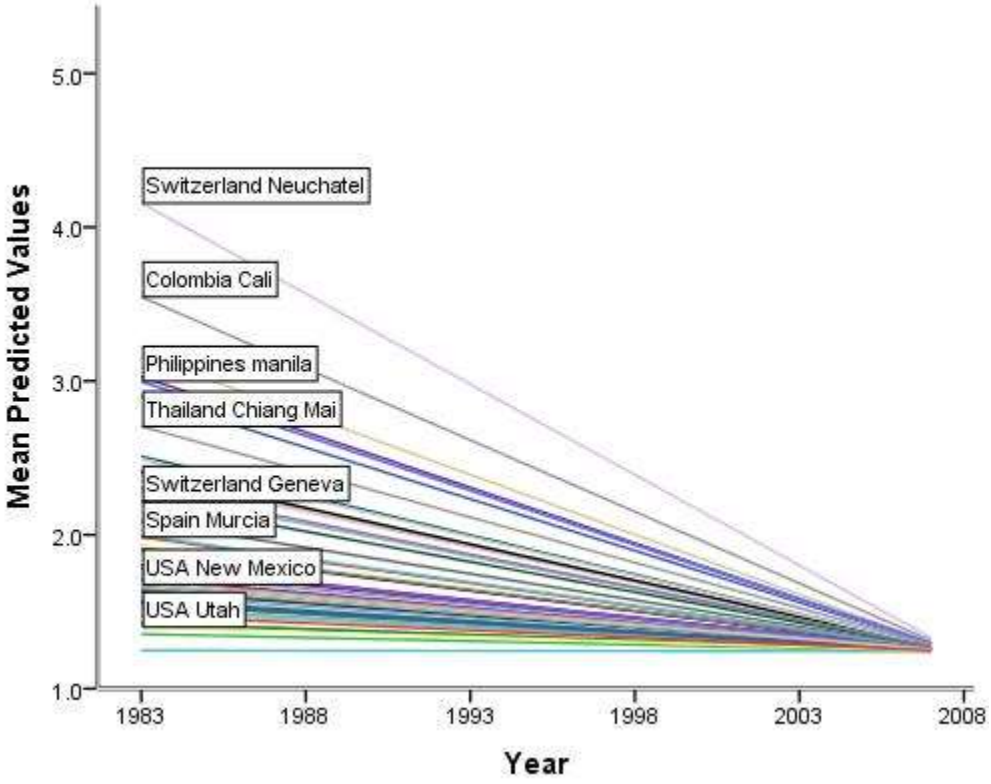
A 7-c Liver cancer: mean predicted value of sex ratios plotted against time 1983-2007.



RI & RS model based on 76 cancer registries where δ^2 (5.4) < τ_1^2 (5.7).

Among cancers that show within-registry population variance that is close to between-component (but still lower) are cancers of pancreas and liver. In liver cancer, this difference is very small (Table 8-14: 5.40 vs 5.74). The analysis of liver cancer also shows that although there was a decrease in the average annual rate of SR, this reduction did not show any significant relationship with passage of time from 1983 to 2007 (β_1 = -0.009: 95% CI: -0.025, 0.006). Figure above shows that the trajectories of lung cancer did not vary significantly (Table 7-14) and that there is very little covariance between these trajectories. The β_0 shows that for liver cancer SR on average at baseline was 5.09, i.e., it is a highly male dominant cancer and the trend is consistent across all registries

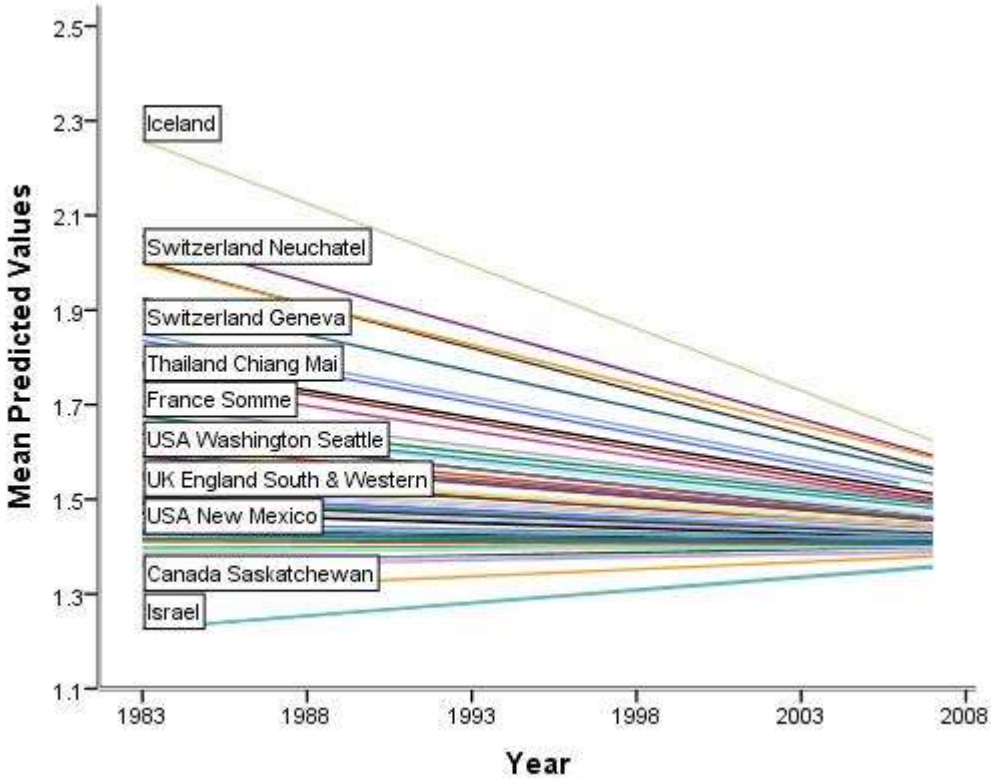
A 7-d Hodgkin's lymphoma: mean predicted value of sex ratios plotted against time 1983-2007.



RI & RS model based on 76 cancer registries where δ^2 (1.57) < τ_1^2 (2.62).

Among other cancers that showed somewhat similar pattern of differences in within- and between- variances of registries for SR, are cancers of bladder, colon and Hodgkin's lymphoma. The baseline SR on average for Hodgkin's lymphoma was 2.74 and average annual rate of decrease i.e., β_1 was – 0.027 (95% CI: -0.038, -0.017). The intercepts, β_0, and trajectories negatively and significantly covaried [Cov (u_{1i} , u_{2j}) τ_{12} = -0.05]. For Hodgkin's lymphoma, the highest baseline SR was from Neuchatel, Switzerland followed by Cali (Colombia), Manila (Philippines) and Chiang Mai in Thailand. One of the lowest baseline SR was found in Utah, US, and was fairly stable throughout the years.

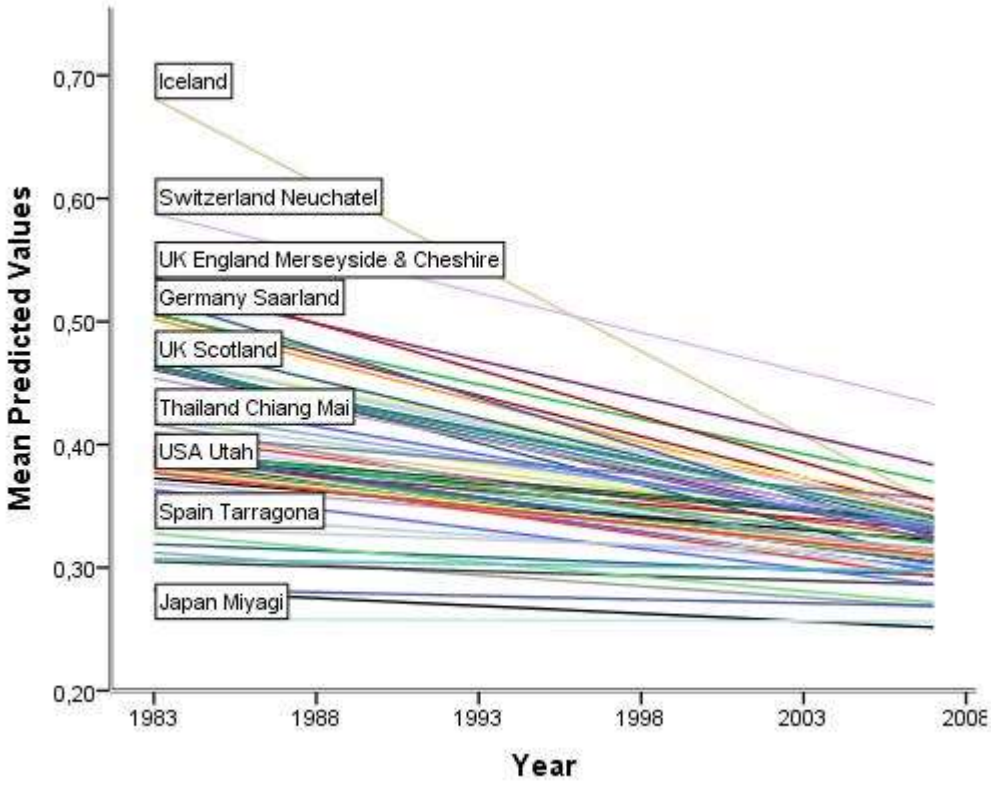
A 7-e NHLymphoma: mean predicted value of sex ratios plotted against time 1983-2007.



RI & RS model based on 76 cancer registries where δ^2 (0.17) $\approx \tau_1^2$ (0.21).

Non-Hodgkin's lymphoma is one of the few cancers along with cancers of thyroid and rectum and anus that showed same within- and between- registry variances of SR. The baseline SR on average for Non-Hodgkin's lymphoma was 1.76, a relatively male dominant but seems to affect females to more or less same extent. The average annual rate of decrease i.e., β_1 was – 0.006 (95% CI: -0.009, -0.002). The intercepts, β_0, and trajectories negatively and significantly covaried [τ_{12} = -0.05]. The highest baseline SR was observed in Iceland whereas lowest was observed in Israel.

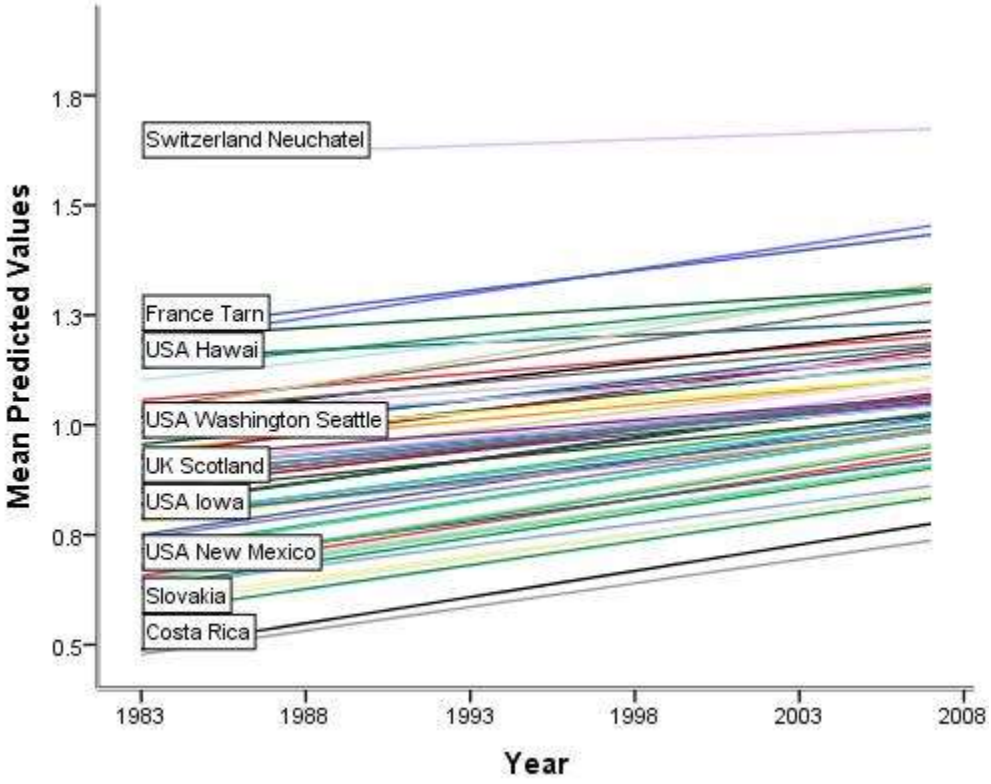
A 7-f Thyroid cancer: mean predicted value of sex ratios plotted against time 1983-2007:



RI & RS model based on 76 cancer registries where δ^2 (0.04) ≈ τ_1^2 (0.03).

Thyroid cancer is highly female dominant cancer as evident from baseline SR of 0.54 and has remained consistently female dominant. It is also one of the few cancers that showed same within- and between- registry variances of SR. The average annual rate of decrease i.e., β_1 was – 0.004 (95% CI: -0.005, -0.003). The intercepts, β_0, and trajectories negatively and significantly covaried [τ_{12} = -0.0005]. The highest baseline SR in 1983 on average was observed in Iceland whereas lowest was observed in Miyagi in Japan.

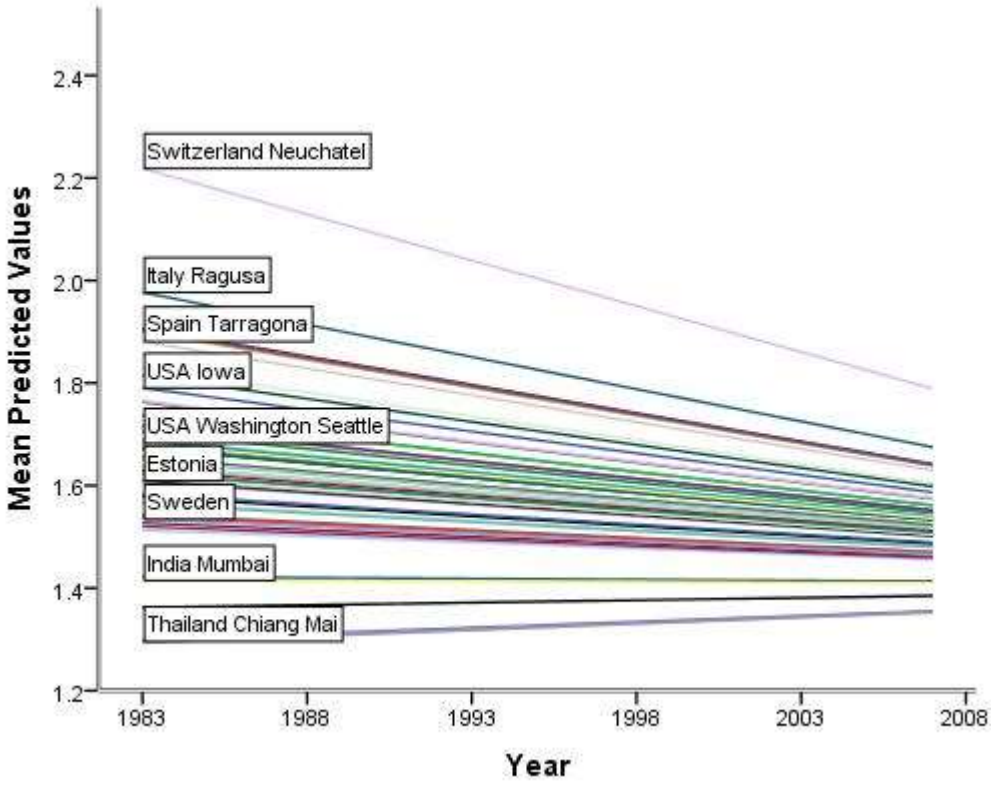
A 7-g Gallbladder cancer: mean predicted value of sex ratios plotted against time 1983-2007.



RI & RS model based on 76 cancer registries where δ^2 (0.38) > τ_1^2 (0.16).

After thyroid, cancer of gallbladder also remained a female dominant cancer, i.e., it showed consistent lowest SR throughout 25 year period. This type of cancer shows within- registry variance that is slightly more than between- registry variance of SR. On average, the baseline SR is 0.65 and the average annual rate of decrease is -0.004. The intercepts, β_0, and trajectories negatively and significantly covaried [τ_{12} = -0.002]. The highest baseline SR in 1983 on average was observed in Switzerland and lowest was observed in Costa Rica.

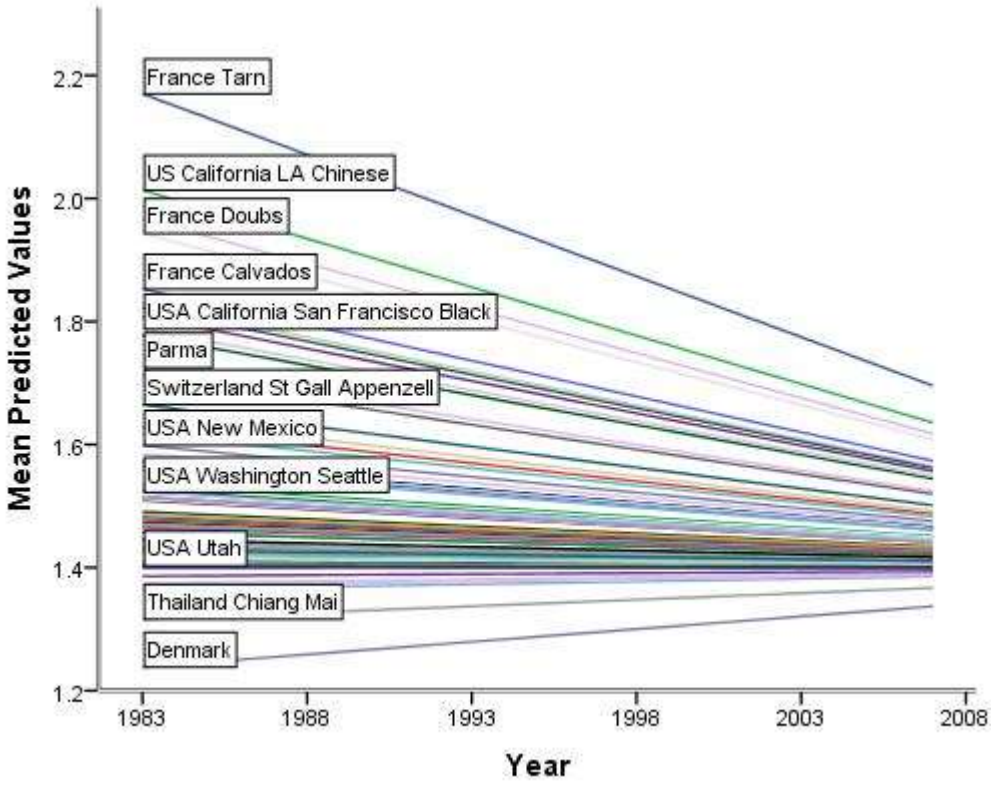
A 7-h Leukemia: mean predicted value of sex ratios plotted against time 1983-2007.



RI & RS model based on 76 cancer registries where δ^2 (0.24) > τ_1^2 (0.12).

Leukemia is considered to be the most stable cancer in terms of incidence rates throughout the world, in particular among children (ref). On average, the baseline SR is 1.85 in our study. The average annual rate of decrease is -0.006 (95% CI: -0.009, -0.002). Leukemia has also one the lowest difference in within- and between- registry variations as seen in Table 8-14 (within- is slightly more than between: 0.24 vs. 0.12). The highest baseline SR in 1983 on average was observed in Neuchatel, Switzerland and lowest was observed in Chiang Mai in Thailand.

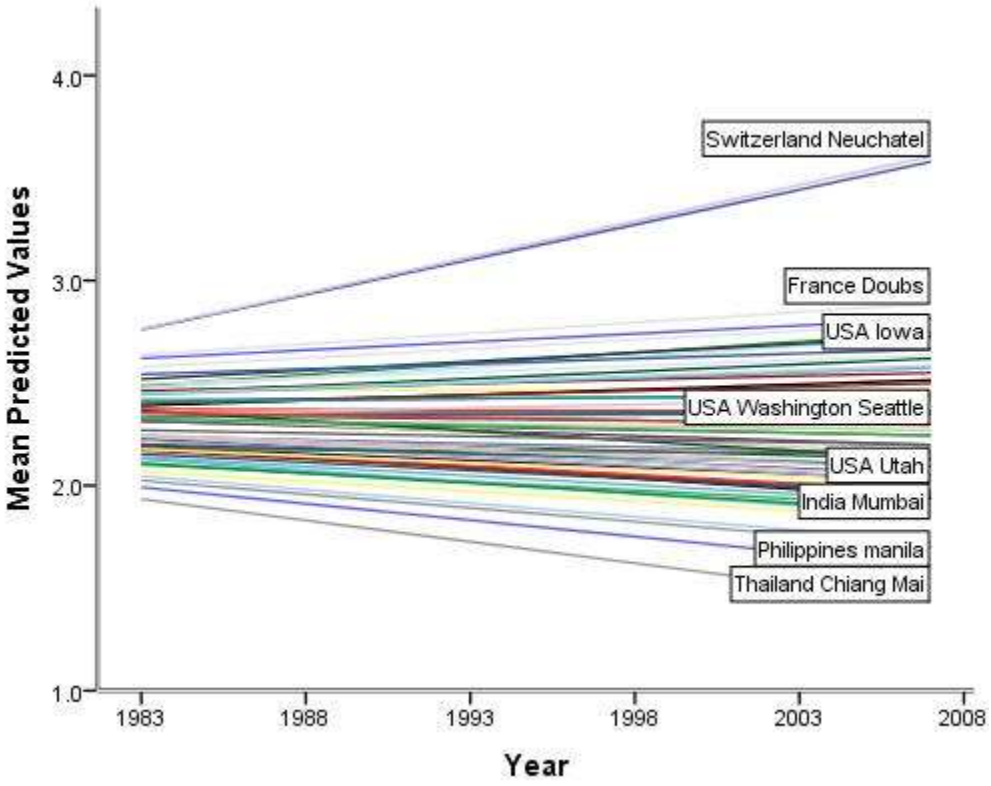
A 7-i Brain cancer: mean predicted value of sex ratios plotted against time 1983-2007.



RI & RS model based on 76 cancer registries where δ^2 (0.17) > τ_1^2 (0.16).

Brain cancer is one of the rare cancer that shows a higher within-registry variations compared to between- registry variations (Table 8-14). The average SR on baseline year of 1983 is 1.65 and somewhat belongs to an intermediate category of SR i.e., neither very highly male dominant nor has very low SR. The average annual rate of decrease is 0.004 (7-14). Tran, France shows one of the steepest reduction in average annual rate of SR from the baseline value of SR in 1983. Registries from Denmark and Thailand show the lowest baseline SR and show a slight increase in average rate of change.

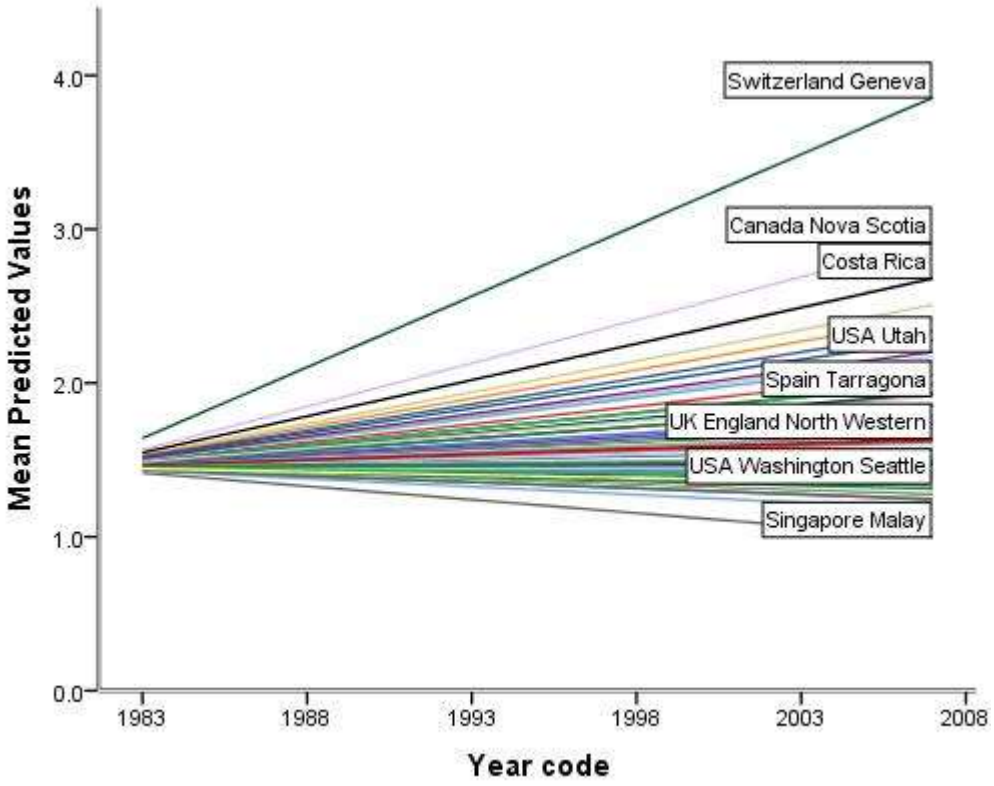
A 7-j Stomach cancer: mean predicted value of sex ratios plotted against time 1983-2007.



RI & RS model based on 76 cancer registries where δ^2 (0.44) > τ_1^2 (0.10).

Although Stomach has a high incidence and the SR on average at baseline is also high (β_0 = 2.74), it is distinct in the sense that the difference between within- and between- registry variance is somewhat similar to brain and gallbladder cancer. The average annual rate of decrease is -0.002 that is not significant (95% CI: -0.007, 0.003). The intercepts, β_0, and trajectories did not show any covarance. The highest baseline SR in 1983 on average was observed in Switzerland and among the lowest were regions belonging to registries from India, Philippines and Thailand.

A 7-h Eye cancer: mean predicted value of sex ratios plotted against time 1983-2007.



RI & RS model based on 76 cancer registries where δ^2 (2.33) > τ_1^2 (0.57).

Cancer of eye is one of the rarest cancers in our analysis. It is one of those cancers where within-registry variation is notably higher than between-registry variation. The SR on average seems to be relatively small indicating that the incidence rates of this cancer is relatively same in both males and females. This type of cancer is also one of those where the average annual rate of SR is increasing however statistically is non-significant (Table 8-14). Cancer of eye also shows that populations with lowest SR have the steepest increase in average annual rate, e.g., Switzerland, Nova Scotia in Canada, and Costa Rica.

Appendix 8: SPSS syntax for categories of sex ratio magnitude (SRm), sex ratio variation (SRv) and Incidence (Im)

Quartiles of SRm (MeanSR), SRv (VarianceSR) and Im (Mean_ASR_MF) and division into 3 equal parts

```
FREQUENCIES VARIABLES=MeanSR  VarianceSR Mean_ASR_MF
  /NTILES=4
  /NTILES=3
  /ORDER=ANALYSIS.
```

Selection of cut-off values in 2003-07 (CI5 Vol X): 3 equal categories [for low (0), medium (1) and high (2)]

```
RECODE MeanSR (0.329 thru 1.435=0) (1.440 thru 2.148=1) (ELSE=2) INTO SRm_cat.
VARIABLE LABELS  SRm_cat 'SRm_cat'.
EXECUTE.

RECODE VarianceSR (0.014 thru 0.173=0) (0.174 thru 1.549=1) (ELSE=2) INTO SRv_cat.
VARIABLE LABELS  SRm_cat 'SRm_cat'.
EXECUTE.

RECODE Mean_ASR_MF (0.208 thru 1.628=0) (1.685 thru 6.516=1) (ELSE=2) INTO Im_cat.
VARIABLE LABELS  SRm_cat 'SRm_cat'.
EXECUTE.
```

Selection of cut-off values in 1988-93 (CI5 Vol VII): 3 equal categories [for low (0), medium (1) and high (2)]

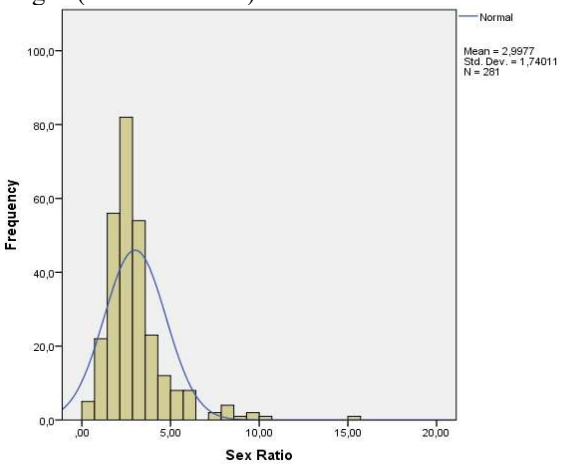
```
RECODE MeanSR (0.376 thru 1.493=0) (1.500 thru 2.230=1) (2.570 thru 15.300=2) INTO SRm_cat.
VARIABLE LABELS  SRm_cat 'SRm_cat'.
EXECUTE.

RECODE VarianceSR (Lowest thru 0.300=0) (0.310 thru 1.240=1) (ELSE=2) INTO SRv_cat.
VARIABLE LABELS  SRv_cat 'SRv_cat'.
EXECUTE.

RECODE Mean_ASR_MF (Lowest thru 1.900=0) (1.950 thru 5.900=1) (ELSE=2) INTO Im_cat.
VARIABLE LABELS  SRm_cat 'SRm_cat'.
EXECUTE.
```

Selection of cut-off values in 1973-77 (CI5 Vol IV): 3 equal categories [for low (0), medium (1) and high (2)]

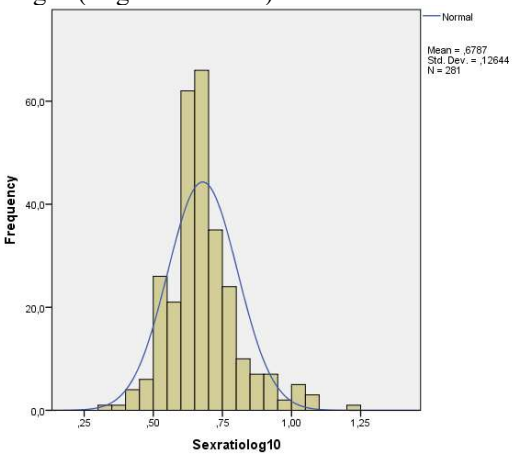
```
RECODE MeanSR (0.425 thru 1.500=0) (1.511 thru 2.030=1) (ELSE=2) INTO SRm_cat.
VARIABLE LABELS  SRm_cat 'SRm_cat'.
EXECUTE.

RECODE VarianceSR (Lowest thru 0.380=0) (0.390 thru 1.360=1) (ELSE=2) INTO SRv_cat.
VARIABLE LABELS  SRv_cat 'SRv_cat'.
EXECUTE.

RECODE Mean_ASR_MF (0.440 thru 2.105=0) (2.110 thru 4.450=1) (ELSE=2) INTO Im_cat.
VARIABLE LABELS  SRm_cat 'SRm_cat'.
EXECUTE.
```

Appendix 9: Distribution of data for selected group of cancers showing log transformations (on right-side)

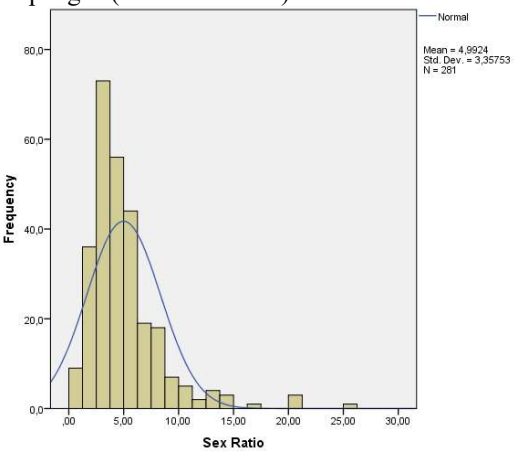
Tongue (Untransformed)



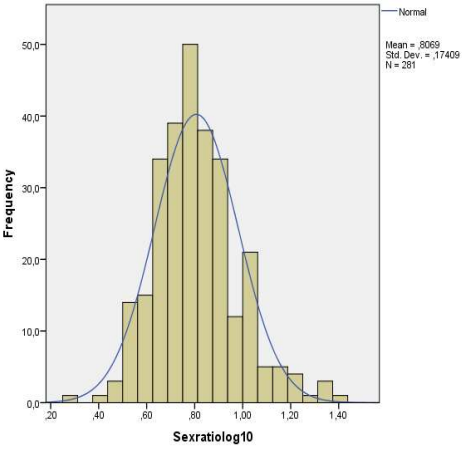
Tongue (Log transformed)



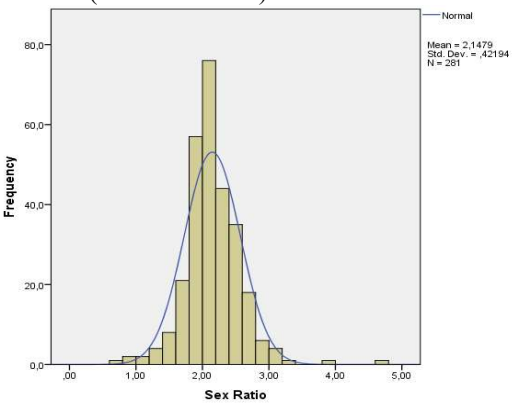
Esophagus (Untransformed)



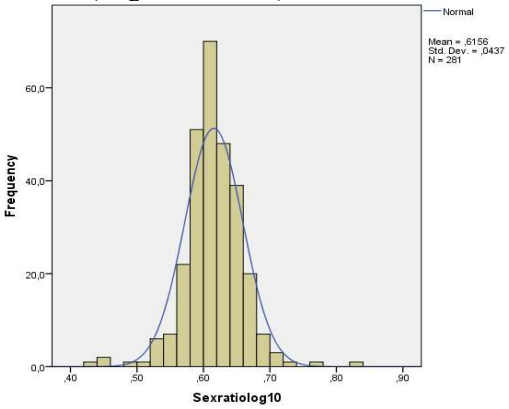
Esophagus (Log transformed)



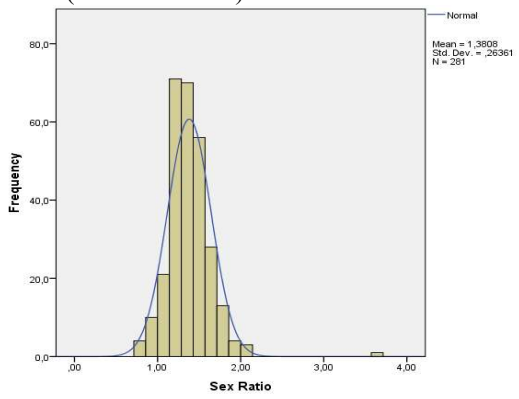
Stomach (Untransformed)



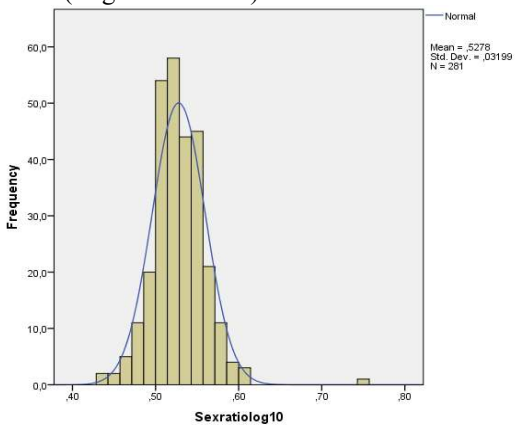
Stomach (Log transformed)



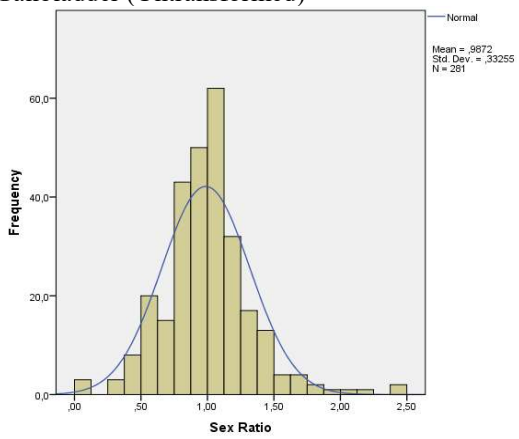
Colon (Untransformed)



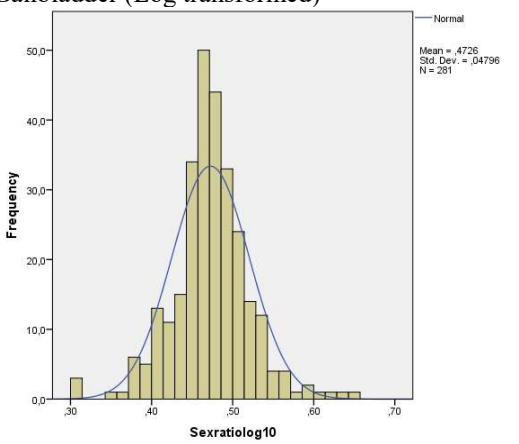
Colon (Log transformed)



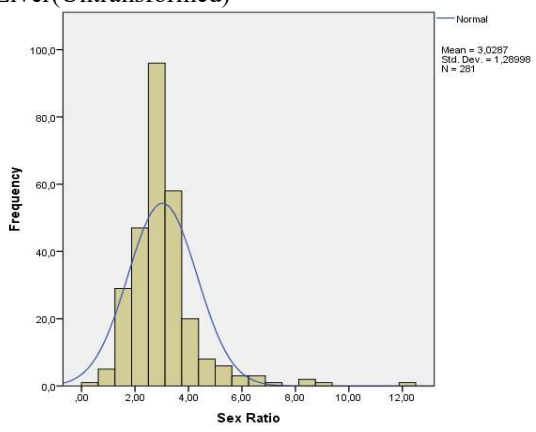
Gallbladder (Untransformed)



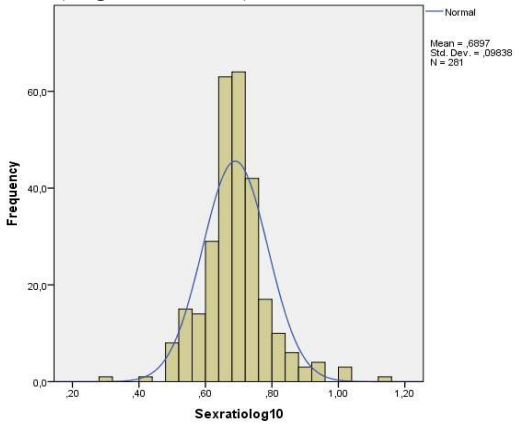
Gallbladder (Log transformed)



Liver(Untransformed)



Liver (Log transformed)



SPSS syntaxes for mixed effect regression models

File name: Syntax_MEM_for all cancers.sps

To select larynx, delete all other cancers with syntax below

```
FILTER OFF.
USE ALL.
SELECT IF (Cancercode = 1).
EXECUTE.
```

A new file Larynx_rough.sav is made with 3158 rows.

ANALYSIS OF 1953-2007

FIXED EFFECT MODEL 1953-2007 with only time (YEARCODE) as predictor

```
UNIANOVA SR WITH Yearcode
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /PRINT=PARAMETER
  /CRITERIA=ALPHA(.05)
  /DESIGN=Yearcode.
```

GRAPH FOR FIXED EFFECT MODEL 1953-2007 (scatterplot: x-axis-yearcode, y-axis-SR, set-color-Registry)

```
* Chart Builder.
GGRAPH
  /GRAPHDATASET NAME="graphdataset" VARIABLES=Yearcode SR Registry MISSING=LISTWISE
REPORTMISSING=NO
  /GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
  SOURCE: s=userSource(id("graphdataset"))
  DATA: Yearcode=col(source(s), name("Yearcode"))
  DATA: SR=col(source(s), name("SR"))
  DATA: Registry=col(source(s), name("Registry"), unit.category())
  GUIDE: axis(dim(1), label("Year code"))
  GUIDE: axis(dim(2), label("Sex Ratio"))
  GUIDE: legend(aesthetic(aesthetic.color.exterior), label("Registry"))
  SCALE: cat(aesthetic(aesthetic.color.exterior), include("0", "1", "2", "3", "4", "5", "6",
    "7", "8", "9", "10", "11"
, "12", "13", "14", "15", "16", "17", "18", "19", "20", "21", "22", "23", "24", "25"
, "26", "27", "28", "29", "30", "31", "32", "33", "34", "35", "36", "37", "38", "39"
, "40", "41", "42", "43", "44", "45", "46", "47", "48", "49", "50", "51", "52", "53"
, "54", "55", "56", "57", "58", "59", "60", "61", "62", "63", "64", "65", "66", "67"
, "68", "69", "70", "71", "72", "73", "74", "75", "76", "77", "78", "79", "80", "81"
, "82", "83", "84", "85", "86", "87", "88", "89", "90", "91", "92", "93", "94", "95"
, "96", "97", "98", "99", "100", "101", "102", "103", "104", "105", "106", "107", "108"
, "109", "110", "111", "112"))
  ELEMENT: point(position(Yearcode*SR), color.exterior(Registry))
END GPL.
```

FIXED EFFECT MODEL 1953-2007 with time (YEARCODE) as predictor and REGISTRY as covariate

```
UNIANOVA SR BY Registry WITH Yearcode
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /PRINT=PARAMETER
  /CRITERIA=ALPHA(.05)
  /DESIGN=Yearcode Registry.
```

RANDOM INTERCEPT MODEL 1953-2007

```
MIXED SR WITH Yearcode
  /CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.000000000001) HCONVERGE(0,
    ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
  /FIXED=Yearcode | SSTYPE(3)
  /METHOD=ML
  /PRINT=SOLUTION TESTCOV
  /RANDOM=INTERCEPT | SUBJECT(Registry) COVTYPE(VC)
  /SAVE=PRED.
```

GRAPH FOR RANDOM INTERCEPT MODEL 1953-2007

```
GGRAPH
  /GRAPHDATASET NAME="graphdataset" VARIABLES=Yearcode MEAN(PRED_1)[name="MEAN_PRED_1"] Registry
    MISSING=LISTWISE REPORTMISSING=NO
  /GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
  SOURCE: s=userSource(id("graphdataset"))
  DATA: Yearcode=col(source(s), name("Yearcode"))
  DATA: MEAN_PRED_1=col(source(s), name("MEAN_PRED_1"))
  DATA: Registry=col(source(s), name("Registry"), unit.category())
  GUIDE: axis(dim(1), label("Year code"))
  GUIDE: axis(dim(2), label("Mean Predicted Values"))
  GUIDE: legend(aesthetic(aesthetic.color.interior), label("Registry"))
  SCALE: cat(aesthetic(aesthetic.color.interior), include("0", "1", "2", "3", "4", "5", "6",
    "7", "8", "9", "10", "11"
, "12", "13", "14", "15", "16", "17", "18", "19", "20", "21", "22", "23", "24", "25"
, "26", "27", "28", "29", "30", "31", "32", "33", "34", "35", "36", "37", "38", "39"
, "40", "41", "42", "43", "44", "45", "46", "47", "48", "49", "50", "51", "52", "53"
, "54", "55", "56", "57", "58", "59", "60", "61", "62", "63", "64", "65", "66", "67"
, "68", "69", "70", "71", "72", "73", "74", "75", "76", "77", "78", "79", "80", "81"
, "82", "83", "84", "85", "86", "87", "88", "89", "90", "91", "92", "93", "94", "95"
, "96", "97", "98", "99", "100", "101", "102", "103", "104", "105", "106", "107", "108"
, "109", "110", "111", "112"))
  ELEMENT: line(position(Yearcode*MEAN_PRED_1), color.interior(Registry), missing.wings())
END GPL.
```

RANDOM INTERCEPT AND SLOPE MODEL 1953-2007

```
MIXED SR WITH Yearcode
  /CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.000000000001) HCONVERGE(0,
    ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
```

```
 /FIXED=Yearcode | SSTYPE(3)
 /METHOD=ML
 /PRINT=SOLUTION TESTCOV
 /RANDOM=INTERCEPT Yearcode | SUBJECT(Registry) COVTYPE(UN)
 /SAVE=PRED.
```

GRAPH FOR RANDOM INTERCEPT AND SLOPE MODEL 1953-2007

```
GGRAPH
 /GRAPHDATASET NAME="graphdataset" VARIABLES=Yearcode MEAN(PRED_2)[name="MEAN_PRED_2"] Registry
   MISSING=LISTWISE REPORTMISSING=NO
 /GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
 SOURCE: s=userSource(id("graphdataset"))
 DATA: Yearcode=col(source(s), name("Yearcode"))
 DATA: MEAN_PRED_2=col(source(s), name("MEAN_PRED_2"))
 DATA: Registry=col(source(s), name("Registry"), unit.category())
 GUIDE: axis(dim(1), label("Year code"))
 GUIDE: axis(dim(2), label("Mean Predicted Values"))
 GUIDE: legend(aesthetic(aesthetic.color.interior), label("Registry"))
 SCALE: cat(aesthetic(aesthetic.color.interior), include("0", "1", "2", "3", "4", "5", "6",
   "7", "8", "9", "10", "11"
, "12", "13", "14", "15", "16", "17", "18", "19", "20", "21", "22", "23", "24", "25"
, "26", "27", "28", "29", "30", "31", "32", "33", "34", "35", "36", "37", "38", "39"
, "40", "41", "42", "43", "44", "45", "46", "47", "48", "49", "50", "51", "52", "53"
, "54", "55", "56", "57", "58", "59", "60", "61", "62", "63", "64", "65", "66", "67"
, "68", "69", "70", "71", "72", "73", "74", "75", "76", "77", "78", "79", "80", "81"
, "82", "83", "84", "85", "86", "87", "88", "89", "90", "91", "92", "93", "94", "95"
, "96", "97", "98", "99", "100", "101", "102", "103", "104", "105", "106", "107", "108"
, "109", "110", "111", "112"))
 ELEMENT: line(position(Yearcode*MEAN_PRED_2), color.interior(Registry), missing.wings())
END GPL.
```

ANALYSIS OF 1983-2007 UNBALANCED DATA

FOR MAKING File with unbalanced data from 1983-2007 Cancer_rough_1983_unbalnced.sav (e.g., Larynx_rough_1983_unbalanced.sav) made from Cancer_rough.sav (e.g., Larynx_rough.sav)

Run the frequency of populations again to check that no population has less than 15 years of follow-up except Russia which has 14 years of followup (freuencies of populations are actually years of followup)

```
DATASET ACTIVATE DataSet2.
FREQUENCIES VARIABLES=Population
 /ORDER=ANALYSIS.
```

For selecting cases from 1983 – 2007 UNBALANCED (all cases before 1983 are de-selected)

```
USE ALL.
COMPUTE filter_$=(Yearcode >= 30).
VARIABLE LABELS filter_$ 'Yearcode >= 30 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
```

```
EXECUTE.
```

For selecting cases from 1983 – 2007 UNBALANCED (all cases before 1983 are deleted)

```
FILTER OFF.
USE ALL.
SELECT IF (Yearcode >= 30).
EXECUTE.
```

A new file is made with unbalanced data from 1983-2007 Cancer_rough_1983_unbalanced.sav (e.g., Larynx_rough_1983_unbalanced.sav) made from Cancer_rough.sav (e.g., Larynx_rough.sav)

FIXED EFFECT MODEL 1983-2007 UNBALANCED with only time (YEARCODE) as predictor

```
UNIANOVA SR WITH Yearcode
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /PRINT=PARAMETER
  /CRITERIA=ALPHA(.05)
  /DESIGN=Yearcode.
```

GRAPH FOR FIXED EFFECT MODEL 1983-2007 UNBALANCED (scatterplot: x-axis-yearcode, y-axis-SR, set-color-Registry)

```
* Chart Builder.
GGRAPH
  /GRAPHDATASET NAME="graphdataset" VARIABLES=Yearcode SR Registry MISSING=LISTWISE
REPORTMISSING=NO
  /GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
  SOURCE: s=userSource(id("graphdataset"))
  DATA: Yearcode=col(source(s), name("Yearcode"))
  DATA: SR=col(source(s), name("SR"))
  DATA: Registry=col(source(s), name("Registry"), unit.category())
  GUIDE: axis(dim(1), label("Year code"))
  GUIDE: axis(dim(2), label("Sex Ratio"))
  GUIDE: legend(aesthetic(aesthetic.color.exterior), label("Registry"))
  SCALE: cat(aesthetic(aesthetic.color.exterior), include("0", "1", "2", "3", "4", "5", "6",
    "7", "8", "9", "10", "11"
, "12", "13", "14", "15", "16", "17", "18", "19", "20", "21", "22", "23", "24", "25"
, "26", "27", "28", "29", "30", "31", "32", "33", "34", "35", "36", "37", "38", "39"
, "40", "41", "42", "43", "44", "45", "46", "47", "48", "49", "50", "51", "52", "53"
, "54", "55", "56", "57", "58", "59", "60", "61", "62", "63", "64", "65", "66", "67"
, "68", "69", "70", "71", "72", "73", "74", "75", "76", "77", "78", "79", "80", "81"
, "82", "83", "84", "85", "86", "87", "88", "89", "90", "91", "92", "93", "94", "95"
, "96", "97", "98", "99", "100", "101", "102", "103", "104", "105", "106", "107", "108"
, "109", "110", "111", "112"))
  ELEMENT: point(position(Yearcode*SR), color.exterior(Registry))
END GPL.
```

FIXED EFFECT MODEL 1983-2007 UNBALANCED with time (YEARCODE) as predictor and REGISTRY as covariate

```
UNIANOVA SR BY Registry WITH Yearcode
```

```
 /METHOD=SSTYPE(3)
 /INTERCEPT=INCLUDE
 /PRINT=PARAMETER
 /CRITERIA=ALPHA(.05)
 /DESIGN=Yearcode Registry.
```

RANDOM INTERCEPT MODEL 1983-2007 UNBALANCED

```
MIXED SR WITH Yearcode
 /CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.000000000001) HCONVERGE(0,
   ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
 /FIXED=Yearcode | SSTYPE(3)
 /METHOD=ML
 /PRINT=SOLUTION TESTCOV
 /RANDOM=INTERCEPT | SUBJECT(Registry) COVTYPE(VC)
 /SAVE=PRED.
```

GRAPH FOR RANDOM INTERCEPT MODEL 1983-2007 UNBALANCED (MEAN PRED_1 on Y asis and Year code on x axis)

```
GGRAPH
 /GRAPHDATASET NAME="graphdataset" VARIABLES=Yearcode MEAN(PRED_1)[name="MEAN_PRED_1"] Registry
   MISSING=LISTWISE REPORTMISSING=NO
 /GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
 SOURCE: s=userSource(id("graphdataset"))
 DATA: Yearcode=col(source(s), name("Yearcode"))
 DATA: MEAN_PRED_1=col(source(s), name("MEAN_PRED_1"))
 DATA: Registry=col(source(s), name("Registry"), unit.category())
 GUIDE: axis(dim(1), label("Year code"))
 GUIDE: axis(dim(2), label("Mean Predicted Values"))
 GUIDE: legend(aesthetic(aesthetic.color.interior), label("Registry"))
 SCALE: cat(aesthetic(aesthetic.color.interior), include("0", "1", "2", "3", "4", "5", "6",
   "7", "8", "9", "10", "11"
, "12", "13", "14", "15", "16", "17", "18", "19", "20", "21", "22", "23", "24", "25"
, "26", "27", "28", "29", "30", "31", "32", "33", "34", "35", "36", "37", "38", "39"
, "40", "41", "42", "43", "44", "45", "46", "47", "48", "49", "50", "51", "52", "53"
, "54", "55", "56", "57", "58", "59", "60", "61", "62", "63", "64", "65", "66", "67"
, "68", "69", "70", "71", "72", "73", "74", "75", "76", "77", "78", "79", "80", "81"
, "82", "83", "84", "85", "86", "87", "88", "89", "90", "91", "92", "93", "94", "95"
, "96", "97", "98", "99", "100", "101", "102", "103", "104", "105", "106", "107", "108"
, "109", "110", "111", "112"))
 ELEMENT: line(position(Yearcode*MEAN_PRED_1), color.interior(Registry), missing.wings())
END GPL.
```

RANDOM INTERCEPT AND SLOPE MODEL 1983-2007 UNBALANCED

```
MIXED SR WITH Yearcode
 /CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.000000000001) HCONVERGE(0,
   ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
 /FIXED=Yearcode | SSTYPE(3)
 /METHOD=ML
 /PRINT=SOLUTION TESTCOV
```

```
 /RANDOM=INTERCEPT Yearcode | SUBJECT(Registry) COVTYPE(UN)
 /SAVE=PRED.
```

GRAPH FOR RANDOM INTERCEPT AND SLOPE MODEL 1983-2007 UNBALANCED (MEAN PRED_2 on Y asis and Year code on x axis)

```
GGRAPH
 /GRAPHDATASET NAME="graphdataset" VARIABLES=Yearcode MEAN(PRED_2)[name="MEAN_PRED_2"] Registry
   MISSING=LISTWISE REPORTMISSING=NO
 /GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
 SOURCE: s=userSource(id("graphdataset"))
 DATA: Yearcode=col(source(s), name("Yearcode"))
 DATA: MEAN_PRED_2=col(source(s), name("MEAN_PRED_2"))
 DATA: Registry=col(source(s), name("Registry"), unit.category())
 GUIDE: axis(dim(1), label("Year code"))
 GUIDE: axis(dim(2), label("Mean Predicted Values"))
 GUIDE: legend(aesthetic(aesthetic.color.interior), label("Registry"))
 SCALE: cat(aesthetic(aesthetic.color.interior), include("0", "1", "2", "3", "4", "5", "6",
   "7", "8", "9", "10", "11"
, "12", "13", "14", "15", "16", "17", "18", "19", "20", "21", "22", "23", "24", "25"
, "26", "27", "28", "29", "30", "31", "32", "33", "34", "35", "36", "37", "38", "39"
, "40", "41", "42", "43", "44", "45", "46", "47", "48", "49", "50", "51", "52", "53"
, "54", "55", "56", "57", "58", "59", "60", "61", "62", "63", "64", "65", "66", "67"
, "68", "69", "70", "71", "72", "73", "74", "75", "76", "77", "78", "79", "80", "81"
, "82", "83", "84", "85", "86", "87", "88", "89", "90", "91", "92", "93", "94", "95"
, "96", "97", "98", "99", "100", "101", "102", "103", "104", "105", "106", "107", "108"
, "109", "110", "111", "112"))
 ELEMENT: line(position(Yearcode*MEAN_PRED_2), color.interior(Registry), missing.wings())
END GPL.
```

ANALYSIS BALANCED DATA 1983-2007

Removing registries before 1983 and selecting year with balanced number of countries n=76 i,e, 76 countries fixed in each year)

```
USE ALL.
COMPUTE filter_$=(Registry = 13 | Registry = 14 | Registry = 17 | Registry = 20 | Registry = 26 |
   Registry = 27 | Registry = 35 | Registry = 37 | Registry = 38 | Registry = 40 | Registry = 43 |
   Registry = 44 | Registry = 45 | Registry = 48 | Registry = 49 | Registry = 51 | Registry = 56 |
   Registry = 61 | Registry = 62 | Registry = 63 | Registry = 64 | Registry = 70 | Registry = 73 |
   Registry = 74 | Registry = 76 | Registry = 77 | Registry = 78 | Registry = 80 | Registry = 81 |
   Registry = 85  | Registry = 88 | Registry = 94 | Registry = 95 | Registry = 107 | Registry = 108 |
   Registry = 110).
VARIABLE LABELS filter_$ 'Registry = 13 | Registry = 14 | Registry = 17 | Registry = 20 | '+
   'Registry = 26 | Registry = 27 | Registry = 35 | Registry = 37 | Registry = 38 | Registry = 40 '+
   '| Registry = 43 | Registry = 44 | Registry = 45 | Registry = 48 | Registry = 49 | R... (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

USE ALL.
```

```
COMPUTE filter_$=(filter_$ = 0).
VARIABLE LABELS filter_$ 'filter_$ = 0 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

FILTER OFF.
USE ALL.
SELECT IF (filter_$ = 1).
EXECUTE.
```

Remove 3 more registries coded 6, 7 and 8 (Australia Western, Austria Tyrol and Austria Vorarlberg)

```
USE ALL.
COMPUTE filter_$=(Registry = 6 | Registry = 7 | Registry = 8).
VARIABLE LABELS filter_$ 'Registry = 6 | Registry = 7 | Registry = 8 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

FILTER OFF.
USE ALL.
SELECT IF (filter_$ = 0).
EXECUTE.
```

Check no of years of followup for every registry which should be 25 years of followup for every registry

```
DATASET ACTIVATE DataSet2.
FREQUENCIES VARIABLES=Population
  /ORDER=ANALYSIS.
```

FIXED EFFECT MODEL 1983-2007 BALANCED with only time (YEARCODE) as predictor

```
UNIANOVA SR WITH Yearcode
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /PRINT=PARAMETER
  /CRITERIA=ALPHA(.05)
  /DESIGN=Yearcode.
```

GRAPH FOR FIXED EFFECT MODEL 1953-2007 (scatterplot: x-axis-yearcode, y-axis-SR, set-color-Registry)

```
* Chart Builder.
GGRAPH
  /GRAPHDATASET NAME="graphdataset" VARIABLES=Yearcode SR Registry MISSING=LISTWISE
REPORTMISSING=NO
  /GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
  SOURCE: s=userSource(id("graphdataset"))
  DATA: Yearcode=col(source(s), name("Yearcode"))
  DATA: SR=col(source(s), name("SR"))
```

```
 DATA: Registry=col(source(s), name("Registry"), unit.category())
 GUIDE: axis(dim(1), label("Year code"))
 GUIDE: axis(dim(2), label("Sex Ratio"))
 GUIDE: legend(aesthetic(aesthetic.color.exterior), label("Registry"))
 SCALE: cat(aesthetic(aesthetic.color.exterior), include("0", "1", "2", "3", "4", "5", "6",
   "7", "8", "9", "10", "11"
, "12", "13", "14", "15", "16", "17", "18", "19", "20", "21", "22", "23", "24", "25"
, "26", "27", "28", "29", "30", "31", "32", "33", "34", "35", "36", "37", "38", "39"
, "40", "41", "42", "43", "44", "45", "46", "47", "48", "49", "50", "51", "52", "53"
, "54", "55", "56", "57", "58", "59", "60", "61", "62", "63", "64", "65", "66", "67"
, "68", "69", "70", "71", "72", "73", "74", "75", "76", "77", "78", "79", "80", "81"
, "82", "83", "84", "85", "86", "87", "88", "89", "90", "91", "92", "93", "94", "95"
, "96", "97", "98", "99", "100", "101", "102", "103", "104", "105", "106", "107", "108"
, "109", "110", "111", "112"))
 ELEMENT: point(position(Yearcode*SR), color.exterior(Registry))
END GPL.
```

FIXED EFFECT MODEL 1983-2007 BALANCED with time (YEARCODE) as predictor and REGISTRY as covariate

```
UNIANOVA SR BY Registry WITH Yearcode
 /METHOD=SSTYPE(3)
 /INTERCEPT=INCLUDE
 /PRINT=PARAMETER
 /CRITERIA=ALPHA(.05)
 /DESIGN=Yearcode Registry.
```

RANDOM INTERCEPT MODEL 1983-2007 BALANCED

```
MIXED SR WITH Yearcode
 /CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.000000000001) HCONVERGE(0,
   ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
 /FIXED=Yearcode | SSTYPE(3)
 /METHOD=ML
 /PRINT=SOLUTION TESTCOV
 /RANDOM=INTERCEPT | SUBJECT(Registry) COVTYPE(VC)
 /SAVE=PRED.
```

GRAPH FOR RANDOM INTERCEPT MODEL 1983-2007 BALANCED

```
GGRAPH
 /GRAPHDATASET NAME="graphdataset" VARIABLES=Yearcode MEAN(PRED_1)[name="MEAN_PRED_1"] Registry
   MISSING=LISTWISE REPORTMISSING=NO
 /GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
 SOURCE: s=userSource(id("graphdataset"))
 DATA: Yearcode=col(source(s), name("Yearcode"))
 DATA: MEAN_PRED_1=col(source(s), name("MEAN_PRED_1"))
 DATA: Registry=col(source(s), name("Registry"), unit.category())
 GUIDE: axis(dim(1), label("Year code"))
 GUIDE: axis(dim(2), label("Mean Predicted Values"))
 GUIDE: legend(aesthetic(aesthetic.color.interior), label("Registry"))
 SCALE: cat(aesthetic(aesthetic.color.interior), include("0", "1", "2", "3", "4", "5", "6",
   "7", "8", "9", "10", "11"
```

```
, "12", "13", "14", "15", "16", "17", "18", "19", "20", "21", "22", "23", "24", "25"
, "26", "27", "28", "29", "30", "31", "32", "33", "34", "35", "36", "37", "38", "39"
, "40", "41", "42", "43", "44", "45", "46", "47", "48", "49", "50", "51", "52", "53"
, "54", "55", "56", "57", "58", "59", "60", "61", "62", "63", "64", "65", "66", "67"
, "68", "69", "70", "71", "72", "73", "74", "75", "76", "77", "78", "79", "80", "81"
, "82", "83", "84", "85", "86", "87", "88", "89", "90", "91", "92", "93", "94", "95"
, "96", "97", "98", "99", "100", "101", "102", "103", "104", "105", "106", "107", "108"
, "109", "110", "111", "112"))
  ELEMENT: line(position(Yearcode*MEAN_PRED_1), color.interior(Registry), missing.wings())
END GPL.
```

RANDOM INTERCEPT AND SLOPE MODEL 1983-2007 BALANCED

```
MIXED SR WITH Yearcode
  /CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.000000000001) HCONVERGE(0,
    ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
  /FIXED=Yearcode | SSTYPE(3)
  /METHOD=ML
  /PRINT=SOLUTION TESTCOV
  /RANDOM=INTERCEPT Yearcode | SUBJECT(Registry) COVTYPE(UN)
  /SAVE=PRED.
```

GRAPH FOR RANDOM INTERCEPT AND SLOPE MODEL 1983-2007 BALANCED

```
GGRAPH
  /GRAPHDATASET NAME="graphdataset" VARIABLES=Yearcode MEAN(PRED_2)[name="MEAN_PRED_2"] Registry
    MISSING=LISTWISE REPORTMISSING=NO
  /GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
  SOURCE: s=userSource(id("graphdataset"))
  DATA: Yearcode=col(source(s), name("Yearcode"))
  DATA: MEAN_PRED_2=col(source(s), name("MEAN_PRED_2"))
  DATA: Registry=col(source(s), name("Registry"), unit.category())
  GUIDE: axis(dim(1), label("Year code"))
  GUIDE: axis(dim(2), label("Mean Predicted Values"))
  GUIDE: legend(aesthetic(aesthetic.color.interior), label("Registry"))
  SCALE: cat(aesthetic(aesthetic.color.interior), include("0", "1", "2", "3", "4", "5", "6",
    "7", "8", "9", "10", "11"
, "12", "13", "14", "15", "16", "17", "18", "19", "20", "21", "22", "23", "24", "25"
, "26", "27", "28", "29", "30", "31", "32", "33", "34", "35", "36", "37", "38", "39"
, "40", "41", "42", "43", "44", "45", "46", "47", "48", "49", "50", "51", "52", "53"
, "54", "55", "56", "57", "58", "59", "60", "61", "62", "63", "64", "65", "66", "67"
, "68", "69", "70", "71", "72", "73", "74", "75", "76", "77", "78", "79", "80", "81"
, "82", "83", "84", "85", "86", "87", "88", "89", "90", "91", "92", "93", "94", "95"
, "96", "97", "98", "99", "100", "101", "102", "103", "104", "105", "106", "107", "108"
, "109", "110", "111", "112"))
  ELEMENT: line(position(Yearcode*MEAN_PRED_2), color.interior(Registry), missing.wings())
END GPL.
```

Cote Scientifique – Fonds de la recherche en santé du Québec

Fonds de la recherche en santé Québec

COTE SCIENTIFIQUE

Code FRSQ : 24D Raza CR 3 ans

PROGRAMME DE BOURSES DE FORMATION DE DOCTORAT

Monsieur Syed Ahsan Raza **Dossier no: 26519**

Au concours du 15 octobre 2011, 493 demandes de bourses de formation de doctorat ont été soumises à l'évaluation de 23 comités d'experts. Les experts de votre comité d'évaluation ont étudié avec attention **les 28** candidatures en compétition. Votre dossier s'est classé **au 2e rang** et vous avez obtenu une cote scientifique de **84.5 sur 100.**

Chaque dossier est évalué sur un total de 100 points. Les évaluateurs doivent respecter les points attribués à chaque critère et sous-critère mais ils ont la liberté de fractionner chaque pointage.

Curriculam Vitae

Syed Ahsan Raza

PROFESSIONAL / QUALIFICATIONS

2010-present **PhD candidate** in Public Health (Epidemiology)
School of Public Health, University of Montreal, Canada.

1999–2001 **MSc** in Epidemiology
Aga Khan University (AKU), Karachi, Pakistan

1990–1996 **MD**
University of Karachi, Sindh Medical College, Karachi, Pakistan

CERTIFICATE COURSES

2016 **Graduate Certificate in Public Health** (Geographic Information Systems) –15 credits
University of North Texas School of Public Health, Texas, United States.

2017 **Short Course in Infectious Disease Modelling** (June 19 to 30, 2017)
London School of Hygiene and Tropical Medicine, London, United Kingdom.

CURRENT STATUS

2016- present **Postdoctoral Fellow – Primary Care Research**
Department of Family and Community Medicine
Baylor College of Medicine, Houston, Texas, USA

FELLOWSHIPS

2009 **International Cancer Technology Transfer Fellowship** (May 1 – October 30 2009)
International Union against Cancer (UICC), Geneva, Switzerland.
Host Institute: International Agency for Research on Cancer (IARC), Lyon, France

2005 - 2007 **Postdoctoral Fellowship – Cancer Epidemiology**
Infections and Cancer Epidemiology Group, Section of Infections
International Agency for Research on Cancer (IARC),
World Health Organization (WHO), Lyon, France

AWARDS

2017 **Best Poster Presentation Award and cash prize**
Title of poster: Worldwide Sex Ratio Variations in Cancer Incidence (PhD Thesis)
By American College of Epidemiology (ACE)
At 2017 Annual Meeting, New Orleans, Louisiana, United States.

2012-2014 **Doctoral Training Scholarship Award,**
Bourses de formation de doctorat
By Quebec Health Research Funds (Fonds de recherche Santé Québec - FRQS),
Government of Québec, Canada.

2011/2012 **Doctoral scholarship competition award (supplement)**
Concours de compléments de bourses – doctorat
By University of Montreal Hospital Research Centre
(Centre de recherché du centre hospitalier de l'université de Montréal – CR-CHUM)

2010-2012 **Scholarship Award of Excellence in PhD-Public Health**
Bourse d'excellence du PhD en santé publique
By School of Public Health, University of Montreal, Canada
(École de santé publique, Université de Montréal, Canada)

CLINICAL TRAINING

1997 Clinical Internship in Pediatric Medicine
National Institute of Child Health, Karachi, Pakistan

1996 Clinical Internship in Surgery
Department of Surgery,
Jinnah Postgraduate Medical Centre, Karachi, Pakistan

1996-2010 Registered Medical Practitioner
Pakistan Medical & Dental Council (PMDC)
Certificate of medical registration no: 30239-S

PREVIOUS ACADEMIC APPOINTMENT

2001-2010 Department of Surgery, Faculty of Health Sciences
Aga Khan University & Hospital, Karachi
Senior Instructor, Research

WORK EXPERIENCE

2004-2010 Manager & Coordinator (Biological Samples Bank & Cancer Research Database, AKU)
2002-2003 Coordinator (Research Unit, Department of Surgery) AKU
2001-2002 Facilitator (Journal Clubs - MSc. Epidemiology & Biostatistics students, AKU)
2000-2001 Teaching Assistant (Epidemiology & Biostatistics for medical students, AKU)
1998-1999 General Physician (Private Medical Practice)

PROFESSIONAL SERVICES

2017-2019 Member, Communications Committee, American College of Epidemiology (ACE)
2017-2019 Member, Steering Committee , Research & Education Foundation, ACE
2007-2010 Member appointed by Dean, Planning committee AKU Cancer Registry

2004-2005 Member, Surgical Grand Round Committee, Dept. of Surgery, AKU
2004-2005 Residents' Budget Committee, Dept. of Surgery, AKU
2004 Member, Scientific Committee, AKU Postgraduate Medical Education Conference.
2002-2004 Member, Quality Assurance Committee, Dept. of Surgery, AKU

PROFESSIONAL MEMBERSHIPS

2017 Life time member, John Snow Society, United Kingdom.
2016 Fellow, Royal Society of Public Health, United Kingdom
2011 Member, Canadian Society for Epidemiology & Biostatistics
2004 Member, American College of Epidemiology

MEDIA INTERVIEW

Dawn News: HPV prevalence among Karachi women low. December 4, 2009.
Available at: http://archives.dawn.com/archives/65479 Cited on several print media in Pakistan
Dawn is the largest English language news network in Pakistan.

INVITED LECTURES

2009 Annual meeting of Society of Obstetricians and Gynaecologists of Pakistan (SOGP)
Human papillomavirus infection and cervical cancer in Pakistan. Karachi, Pakistan.

2008 International Symposium on Tropical Medicine and Hygiene
Title of lecture: 'Worldwide multi-centre HPV prevalence surveys'
Aga Khan University, Karachi, Pakistan

2008 World Cancer Conference,
Title: Epidemiology & Prevention of hepatitis associated hepatocellular carcinoma- a worldwide translational approach, Beijing, China

2005 Research Seminar for Medical students, Postgraduate students and Faculty,
Army Medical College of National University of Science & Technology,
Session title: Study Designs in clinical research, Rawalpindi, Pakistan

2004 Course series in: Nurture of student-faculty interaction.
Aga Khan University, Karachi, Pakistan.

GRANTS

2009-2011 AKU Research Council Grant (seed money grant): US$ 3000
Comparison of analgesic efficacy of intraperitoneal lignocaine with bupivacaine after laparoscopic cholecystectomy: a randomized controlled trial.
Co-Investigator

2007-2008 IARC-Return Home Grant US$ 10,000
Secular trends and geographic variations in hepatitis B and C virus associated hepatocellular carcinoma in Pakistan.
Principal Investigator

2006-2009 Prevalence survey of human papilloma virus infection and cervical neoplasia in Pakistan
US$ 35,000 (Postdoctoral project supervised by Dr. Silvia Franceschi, IARC, WHO)
Principal Investigator

2006-2008 AKU Research Council Grant: US$14,000
Sentinel Lymph node biopsy as a predictor of metastatic disease
in cervical lymph node from cancer of oral cavity
Co-Investigator

2002-2010 AKU Research Council Grant: US$50,000
Surgical Oncology Research Database (Renamed: Biological Samples Bank & Cancer Research Database
Co-Investigator

2002-2008 AKU Research Council Grant: US$19,827
Risk factors for intestinal metaplasia in the gallbladder epithelium of patients with gallstones
Co-Investigator

2000-2001 AKU Research Council Grant: US$3,404 (MSc thesis)
Risk factors for neonatal tetanus in Karachi: matched case control study
Principal Investigator

GRANT REFEREE

Member of Peer Review Committee, Canadian Institute of Health Research (CIHR), Canada.

SUPERVISION OF RESEARCH PROJECTS

Following research projects of medical students, residents and junior faculty were under my supervision/co-supervision and assistance during my tenure from November 2001 to March 2010 as Senior Research Instructor in Department of Surgery, Aga Khan University and Hospital, Karachi, Pakistan:

Impact of age on outcome after colorectal cancer surgery in the elderly.
Hassan Bari (Resident, General Surgery)

Early postoperative outcome after curative colorectal cancer surgery.
Hassan Bari (Resident, General Surgery)

Transfer delay and in-hospital mortality of trauma patients (General Surgery)
Afrasyab Khan (Medical student), Syed Nadir Naeem (Resident, General Surgery)

Risk factors for local recurrence after breast conserving surgery (General Surgery)

Quality of life in Head and Neck Cancer patients
Moghira Iqbal (Resident, Otolaryngology, Head & Neck Surgery).

Predictors of recurrence in oral cancers at a tertiary care hospital in Karachi.

Ali Abbas (Resident, Otolaryngology, Head & Neck Surgery).

Factors affecting mortality in emergency abdominal surgery among elderly patients
Kulsoom Faizullah (Resident, General Surgery)

Predictors of complications for central venous access devices in children < 15 years of age.
Afzal Jatt (Resident, Pediatric Surgery)

Survival patterns of buccal cancer in patients with neck metastasis
Shehzad Ghaffar (Instructor, Otolaryngology, Head & Neck Surgery).

Predictors of recurrence in cancers of cheek.
Shehzad Ghaffar (Instructor, Otolaryngology, Head & Neck Surgery).

Endoscope and middle ear surgery
Shehzad Ghaffar (Instructor, Otolaryngology, Head & Neck Surgery)

Prevalence and factors associated with fecal incontinence among females > 15 years old in Karachi
Turab Pishori (Assistant Professor, General Surgery)

Conversion from laparoscopic to open cholecystectomy
Mohammad Tayyab (Resident, General Surgery)

PUBLICATIONS

Raza SA. Theory of scientific investigation by Hempel and a case of Semmelweis. ***J Family Med Prim Care*** 2017;6(2):198-200.

Clifford GM, Waterboer T, Dondog B, Qiao YL, Kordzaia D, Hammouda D, Keita N, Khodakarami N, **Raza SA**, Sherpa AT, Zatonski W, Pawlita M, Plummer M, Franceschi S. Hepatitis C virus seroprevalence in the general female population of 9 countries in Europe, Asia and Africa. ***Infect Agent Cancer*** 2017 Feb 2;12:9.

Avan BI, **Raza SA**, Kirkwood B. An epidemiological study of urban and rural children in Pakistan: examining the relationship between delayed psychomotor development, low birth weight and postnatal growth failure. ***Trans R Soc Trop Med Hyg***. 2015 Mar;109(3):189-96

Avan BI, **Raza SA**, Kirkwood BR. A community-based study of early childhood sensory stimulation in home environment associated with growth and psychomotor development in Pakistan. ***Int J Public Health*** 2014;59(5):779-88.

Raza SA, Avan BI. Disposable clean birth kits and prevention of neonatal tetanus infections in the presence of skilled delivery attendants. ***Int J Gynaecol Obstet*** 2013;120:148-51.

Khan MR, Raza R, Zafar SN, Shamim F, **Raza SA**, Pal KI, Hasnain Z, Alvi R, Chawla T, Azami R. Intraperitoneal lignocaine (lidocaine) versus bupivacaine after laparoscopic cholecystectomy: Results of a randomized controlled trial. ***J Surg Res*** 2012;178(2):662-9.

Bhatti JA, Akhtar U, **Raza SA**, Ejaz K. Selecting a research topic. ***J Pak Med Assoc*** 2012 62(2):184-6.

Khan MR, Bari H, Zafar SN, **Raza SA**. Impact of age on outcome after colorectal cancer surgery in the elderly - a developing country perspective. ***BMC Surg*** 2011:17;11:17.

Khan MR, **Raza SA**, Ahmad Z, Naeem S, Pervez S, Siddiqui AA, Ahmed M, Azami R. Gallbladder intestinal metaplasia in Pakistani patients with gallstones. ***Int J Surg*** 2011;9(6):482-5.

Khan MR, Bari H, **Raza SA**. Early postoperative outcome after curative colorectal cancer surgery. ***Singapore Med J*** 2011;52(3):195-200.

Raza SA, Franceschi S, Pallardy S, Malik FR, Avan BI, Zafar A, Ali SH, Pervez S, Serajuddaula S, Snijders PJ, van Kemenade FJ, Meijer CJ, Shershah S, Clifford GM. Human papillomavirus infection in women with and without cervical cancer in Karachi, Pakistan. ***Br J Cancer*** 2010 25;102(11):1657-60.

Khan A, Zafar H, Naeem SN, **Raza SA**. Transfer delay and in-hospital mortality of trauma patients in Pakistan. ***Int J Surg*** 2010;8(2):155-8.

Franceschi S, **Raza SA**. Epidemiology and prevention of hepatocellular carcinoma-A worldwide translational approach. ***Cancer Letters*** 2009 Dec 1;286(1):5-8

Raza SA, Clifford GM, Franceschi S. Worldwide variation in the relative importance of hepatitis B and hepatitis C viruses in hepatocellular carcinoma. ***Br J Cancer*** 2007; 96: 1127-1134.

Avan BI, Rahbar MH, **Raza SA**. The role of family configuration in early childhood intellectual development in the context of an extended family system. ***J Postgrad Med*** 2007;53:27-33.

Ghaffar S, Ikram M, Zia S, **Raza A**. Incorporating the endoscope into middle ear surgery. ***Ear Nose Thorat J*** 2006;85(9):593-6

Avan BI, **Raza SA**, Khokhar S, Awan F, Hashmi A, Sohail N, Hamza H. Work environment of residency programs in a developing country setup. ***J Postgrad Med*** 2006;52 (1):11-6; discussion 17-8.

Ahmad K, **Raza SA** Janjua NZ, Hamza H, Khan MI. Interventions for improvement in unsafe injection practices in Pakistan. ***Int J Infect Dis*** 2005;9(4):232-3.

Tayyab M, **Raza SA**, Khan MR, Azami R. Conversion from laparoscopic to open cholecystectomy: multivariate analysis of preoperative risk factors. ***J Postgrad Med*** 2005;51(1):17-20.

Tayyab M, **Raza SA**, Khan MR, Azami R. Author's reply**.** ***J Postgrad Med*** 2005;51(2):153.

Avan BI, **Raza SA**, Afridi H. Residents' perceptions of communication skills in postgraduate medical training programs of Pakistan. ***J Postgrad Med*** 2005;51(2):85-91.

Raza SA, Akhtar S, Avan BI, Hamza H, Rahbar MH. A matched case control study of risk factors for neonatal tetanus in Karachi, Pakistan. ***J Postgrad Med*** 2004;50:247-52.

Avan BI, **Raza SA**, Hamza H, Khokhar S, Awan F. Factors influencing selection of surgical specialty among Pakistani medical graduates. ***J Postgrad Med*** 2003; 49:197-201.

In progress:

Raza SA, Clifford GM, Franceschi S et al. Secular trends and geographic variations in hepatitis B and C virus associated hepatocellular carcinoma in Pakistan.

Book chapter

Raza SA: Community ECD Program Development. In: Early Childhood Development. (Avan BI [editor]) Chapter 11. London, New York, Karachi: Oxford University Press 2007.
Available at: http://www.oup.com/uk/catalogue/?ci=9780195473896

E-letters

Bhatti JA, **Raza SA**, Nasrullah M. The understated aspect in advocating efficient cookstoves. (Reply to: WJ Martin II *et al*., A major Environmental cause of death. ***Science***, October 2011: 180-181.
Available at: http://www.sciencemag.org/content/334/6053/180/reply

Ahmad K, Janjua NZ, Khan MI, Hamza H, **Raza SA**. Will BMJ and its sister journals save some rain forest? *BMJ* 2005 (Rapid Response for Rocha J., Brazil fights to save biodiversity, ***BMJ*** 1994;308:75).
Available at: http://www.bmj.com/rapid-response/2011/10/30/will-bmj-and-its-sister-journals-help-save-some-rainforests

Oral presentations/Abstracts/ Posters

Raza SA. Using geospatial distribution of age-adjusted gender disparities in cancer incidence to explore etiologic hypotheses. Oral presentation at *11th International Symposium on Geospatial Health, Baltimore, Maryland, United States, November 4, 2017*.

Raza SA. Exploring the global impact of gender biases in cancer registration on reported gender disparities in cancer incidence. Poster presented at *2017 Global Health Workshop of American Academy of Family Physicians (AAFP), Houston, Texas, October 4-6, 2017.*

Raza SA, Tahir MR, Schnitzer M, Siemiatycki J. Worldwide Sex Ratio Variations in cancer Incidence. Poster presented *at 2017 Annual Meeting of American College of Epidemiology (ACE), New Orleans, Louisiana, United States, September 24-26, 2017 (Best Poster Award Prize). Abstract published in Annals of Epidemiology*.

Raza SA, Neonatal survival in low resource settings: Are disposable clean delivery kits effective in prevention of neonatal tetanus infections in the presence of skilled delivery attendants?. Poster in *141st American Public Health Association meeting, Boston MA, United States, November 2-6, 2013*.

Raza SA, Franceschi S, Malik FR, Serajuddaula S, Snijders PJF, Meijer CJLM, Shershah S, Clifford GM. Human papillomavirus infection in women with and without cervical cancer in Karachi, Pakistan. Abstract and poster in *26th International Papilloma Conference and Clinical and Public Health Workshop. Montreal, Canada, July 2010.*

Raza SA. Human papillomavirus infection and cervical cancer in Pakistan. Karachi, Pakistan. Oral presentation in *13th Biennial Scientific Conference and annual meeting of Society of Obstetricians and Gynaecologists of Pakistan (SOGP), Karachi, Pakistan, February 5-7, 2009*.

Raza SA. Worldwide multi-centre HPV prevalence surveys. Oral presentation in *International*

Symposium on Tropical Medicine and Hygiene organized by Aga Khan University, Karachi and Royal Society of Tropical Medicine and Hygiene, United Kingdom held at Aga Khan University, Karachi, Pakistan. November 26, 2008.

Raza SA, Franceschi S. Epidemiology & Prevention of hepatitis associated hepatocellular carcinoma- a worldwide translational approach. Oral presentation in *2nd Annual International Congress and World Cancer Conference, Beijing, China, November 20-22, 2008.*

Raza SA, Abbas F, Khan S, Ikram M, Aziz A, Pervaiz S, Chawla T, Khan S, Enam A, Frossard P, Siddiqui A. Biological Samples Bank and Cancer Research Database: Experience from pilot project at Aga Khan University Hospital, Karachi, Pakistan. Poster presented at *Joint Symposium on Integrative Molecular Cancer Epidemiology, International Agency for Research on Cancer (IARC), American Association for Cancer Research (AACR), and European Association for Cancer Research (EACR). Lyon, France, July 2008.*Abstract published in *Eur J Cancer Supplements*, July 2008; (9):199.

Raza SA, Clifford GM, Franceschi S. Worldwide variation in the relative importance of hepatitis B and hepatitis C viruses in hepatocellular carcinoma. Abstract and poster in *Annual meeting of American Association for Cancer Research (AACR), Los Angeles, California, United States, April 2007.*

Raza SA, Clifford GM, Franceschi S. Role of hepatitis B and hepatitis C viruses in hepatocellular carcinoma. Poster presented at *Annual Scientific day, International Agency for Research on Cancer (IARC), Lyon, France, May 2007*.

Abbas A, **Raza SA**, Awan S. Multivariate analysis of predictors of recurrence of oral cancers. Abstract and poster in *Health Sciences Research Assembly, Aga Khan University, Karachi, Pakistan, March 2005*.

Avan BI, **Raza SA**, Hamza H, Khokhar S, Awan F. Factors influencing selection of surgical specialty among Pakistani medical graduates. Abstract and poster in *Health Sciences Research Assembly, Aga Khan University, Karachi, Pakistan, March 2005.*

Faizullah K, **Raza SA**, Zafar H. Emergency abdominal surgery among elderly patients. Abstract and poster presented at *Health Sciences Research Assembly, Aga Khan University, Karachi, Pakistan, March 2005.*

Ghafar S, Ikram M, Zia S, **Raza SA**. Role of endoscope in middle ear surgery-AKUH experience. Abstract and poster in *Health Sciences Research Assembly, Aga Khan University, Karachi, Pakistan, March 2005.*

Tayyab M, **Raza SA**, Khan MR, Azami R. Conversion from laparoscopic to open cholecystectomy: multivariate analysis of preoperative risk factors. Abstract and poster in *Health Sciences Research Assembly, Aga Khan University, Karachi, Pakistan, March 2005.*

Iqbal M, **Raza SA**, Suhail A, Ikram M. Quality of life in Head and Neeck cancer patients. Abstract and poster in *Health Sciences Research Assembly, Aga Khan University, Karachi, Pakistan, March 2005.*

Jat A, Hussain S, **Raza SA**, Hoodbhoy Z. Predictors of complications for central venous access devices in children younger than 15 years of age. Abstract and poster in *Health Sciences Research Assembly, Aga Khan University, Karachi, Pakistan, March 2005.*

Raza SA, Akhtar S, Avan BI, Hamza H, Rahbar MH. A matched case control study of risk factors for neonatal tetanus in Karachi, Pakistan. Abstract and poster in *Health Sciences Research Assembly, Aga Khan University, Karachi, Pakistan, March 2004.*

Khan MR, Tayyab M, Chawla T, **Raza SA**, Azami R. Laparoscopic and open cholecystectomies, a perspective from developing country. Abstract in *J Soc Laparoendosc Surg 2003;7:S80.*

JOURNAL REFEREE

Have refereed for the following journals:

International Journal of Cancer;
International Journal of Surgery;
The Pediatric Infectious Disease Journal;
BMC Medical Education;
Minimally Invasive Surgery.
Cancer Biology and Medicine
Annals of Epidemiology

PARTICIPATION IN SEMINARS/WORKSHOPS

International Society of Geospatial Health Workshop on Spatial Statistics, (course director: Donal Bisanzio), November 3, 2017, Maryland, United States, November 3, 2017.

American College of Epidemiology Minority Affairs Committee Workshop. Meeting of Minds: Why haven't compelling epidemiologic data eliminate health disparities? New Orleans, Louisiana, United States, September 23 2017.

American College of Epidemiology Q-GIS basics workshop (course instructor: Amber Dismer, CDC). New Orleans, Louisiana, United States, September 24, 2017.

Epidemiology Student Conference, Canadian Society for Epidemiology and Biostatistics, Montreal, Canada, June 2011.

Third North American Congress of Epidemiology. Making maps: Exploratory data analysis with GIS. Montreal, Canada, June 2011.

Third North American Congress of Epidemiology. Eco-Epidemiology: The rational and methods for understanding the complex dynamics between ecosystem change and human health. Montreal, Canada, June 2011

IARC Summer School in Cancer Epidemiology. Module 1: Analytical Epidemiology; Module 2: Survival Analysis Methods for Cancer Registries, Lyon, France, June 19-July 14, 2006.

IARC Seminar on Biobanks in Europe, Lyon, France 2006
International Course on Cancer Epidemiology- Principles and Methods – organized by Massey University, New Zealand and Ministry of Health, Kingdom of Tonga (course director: Professor Neil Pearce). Nuku'alofa, Tonga, May 17-28, 2004.

Professional development workshop: Master class in professional life. (by Professor Rosslyne Freeman, AKU), Office of Faculty Development, Aga Khan University. Karachi, Pakistan, October 2003.

Second Symposium and workshop on ‘Ethical issues of health research in developing countries, Aga Khan University. Karachi, Pakistan, August 2003.

Ninth Annual Cancer Conference by Pakistan Society of Clinical Oncology. Karachi, Pakistan, December 2002.

National conference on ‘Early childhood development and care; Best investment for future, Aga Khan University. Karachi, Pakistan, September 2002.

National symposium, Aga Khan University. Complexity and Health Care. Workshop on ‘Making decisions in chaos and understanding chaos in decision making. Karachi, Pakistan, May 2002.

Facilitations and tools training workshop, continuous quality improvement (CQI), Aga Khan University. Karachi, Pakistan, March 2002.