Synopsis of the Ph.D. Thesis

Title: Statistical Modeling and Projections for Breast and Reproductive Tract Cancers in Kerala

Research scholar: Aleyamma Mathew
Research Guide: Dr. K.K.Jose

1. Introduction

Burden of breast and reproductive tract cancers in Kerala is likely to become more acute. Morbidity and mortality rates are routinely used to quantify the burden of cancer, but these are often used separately. The Disability-Adjusted Life Years (DALY) is a summary measure that can reflect the burden of both morbidity and mortality in a single indicator (Murray and Lopez, 1996a; Murray and Lopez, 1996b). DALY is estimated by adding the number of years of life a person loses as a consequence of dying early because of the disease (Years of Life Lost, YLL); and the number of years of life a person lives with disability caused by the disease (Years of Life lived with Disability, YLD). Competing risk scenario arises when death is attributed to several risks unambiguously. Several approaches are available to estimate DALY with or without considering the effect of competing risk.

Estimate of the probability of developing or dying due to cancer, either over a lifetime or over a specified number of years, is another useful summary measure of the burden of cancer in a population. Probabilities of developing or dying due to cancer can be estimated after removing the risk due to other causes. Using competing risk models, the probabilities of death due to various risks can be

separated. Several approaches are available for estimating these probabilities with and without considering the effect of competing risk.

Further, examination of how cancer incidence or mortality rates have been changing over time is of interest for usual questions such as why the risk has been changing, and what is likely to happen in future? Different regression models have been used in time-trend analysis of cancer data to estimate the annual percent change. A more useful answer to the question of what is likely to happen in future trends in cancer risk requires projection of burden cancer in the community.

2. Review of literature

i. Burden of cancer using disability adjusted life years (DALY)

DALY for each cancer-age group is calculated as the sum of the burden of premature mortality i.e. Years of Life Lost (YLL) and Years Lived with Disability (YLD). Components of YLD have been estimated by using the software DISMOD and DALY have been estimated using the templates provided in the Global Burden of Disease (GBD) study conducted by WHO (Murray and Lopez 1996a,b; Murray and Lopez 1997a,b). YLD is estimated by assuming that death due to disease under study and death due to other causes are independent. With this assumption, the disease model can be completely determined by three transition hazards viz. incidence, remission and case-fatality. However, for disease like cancer, information on remission is difficult to obtain as it requires follow-up data. Secondly, deaths due to cancer may not be independent on other causes and thus other cause mortality may also be considered. After removing the risk due to other causes (competing risks), YLD is estimated by using the indicators such as incidence, mortality and all-cause mortality.

Choices such as disability weight, discount rate, age-weighting, duration of disease and age at onset are made for estimating both YLL and YLD. Disability weight reflects the severity of the disease on a scale from “perfect health” to “equivalent to death”. This weight is derived using different methods such as rating scale or visual analogue scale, standard gamble, time trade off and person trade off. Discount rate reflects the rate at which society as a whole is willing to trade off present for future benefits.The value of a life year is set higher than the value of future life years. Discounting future benefits is of standard practice in economic analysis. Age weighting is based on a number of studies that have indicated that there is a broad social preference to value a year lived by a young adult more highly than a year lived by a young child or at older ages (Murray and Lopez 1996a,b). The formula for YLL under the various above social value choices are:

i) Without discounting or age-weights:

$YLL = N * L$; where N is the number of deaths; L standard life expectancy.

ii) With non-zero discounting (at a rate of 3%) and uniform age-weights:

$$YLL = \frac{N}{0.03}\left(1 - e^{-0.03L}\right)$$

iii) With non-zero discounting (at a rate of 3%) and non-zero age-weights:

$$YLL = N * Ce^{(ra)} / (\beta+r)^2 \left[e^{-(\beta+r)(L+a)}\left[-(\beta+r)(L+a)-1\right] - e^{-(\beta+r)a}\left[-(\beta+r)a-1\right]\right]$$

where r is the discount rate, C is the age-weighting correction constant (GBD standard value is 0.1658), β is the parameter from the age-

weighting function (GBD standard value is 0.04), *a* is the age at onset, and *L* standard life expectancy at age *a*.

Using DISMOD II procedure, average duration of disability and age at onset are derived based on either incidence, mortality and case-fatality rate (without considering competing risk approach) or based on incidence, mortality and RR mortality (considered all-cause mortality).

The formula for YLD under the various social value choices are:

*i) YLD = I * DW * L;* where *I* is the number of incident cases in the reference period, *DW* is the disability weight (in the range 0-1) and *L* is the average duration of disability (measured in years).

ii) With non-zero discounting (at a rate of 3%), the formula becomes:

$$YLD = \frac{I * DW * (1 - e^{-0.03L})}{0.03}$$

iii) With non-uniform age weights, YLD is given by:

$$YLD = I*DW*C\, e^{(ra)} / (\beta+r)^2 \, [e^{-(\beta+r)(L+a)} \, [-(\beta+ r)(L+a)-1] - e^{-(\beta+ r)a} \, [-(\beta+r)a-1]]$$

where *DW* is the disability weight, *r* is the discount rate (GBD standard value is 0.03), *C* is the age-weighting correction constant (GBD standard value is 0.1658), β is the parameter from the age-weighting function (GBD standard value is 0.04), *a* is the age of onset, and *L* is the duration of disability.

Using a parameter *K* that specifies whether age weighting is applied (*K*=1) or not (*K*=0), the non-zero discounting and non-uniform age-weighting formulae can be combined into a single general formula for YLD and which is given below.

$$YLD = I*DW\{ K\, Ce^{(ra)} / (\beta+r)^2 \, [e^{-(\beta+r)(L+a)} \, [-(\beta+ r)(L+a)-1] - e^{-(\beta+ r)a} \, [-(\beta+r)a-1]] + (1-K)\, (L/r)\, (1- e^{-rL})\}$$

ii. Probability of developing cancer

An estimator for the age-conditional probability of cancer is derived using statistical competing risk model which allows the calculation of the age-conditional probabilities for any given age-range (Fay et al., 2003). This estimator requires cancer incidence and mortality rates as well as all-cause mortality rates. The age-conditional probabilities are estimated using the software DevCan 6.4.1.

In cancer registries, incidence and mortality rates are available, but all cause mortality may not be available. Using the incidence rates, probability of developing cancer in terms of cumulative risk is estimated assuming that the risk due to other causes is negligible. Overall incidence of a disease observed in a population is described as the cumulative incidence rate which provides an approximation of the risk of developing a disease. Thus cumulative rate over a whole lifetime is an integral function represented by the incidence curve.

iii. Time-trend in incidence rates of cancer

Various analytic approaches and measures of trend, including graphical display and overall or mean annual percentage change in the age-standardized rate or age-specific rates have been used to study trends in incidence of cancers.

Estimated Annual Percent Change (APC) is one way to characterize trends in cancer rates over time (Christensen, 1997). In this method, cancer rates are assumed to change at a constant percentage of the rate of the previous year. Rates that change at a constant percentage every year linearly on a log scale (log-linear model or Poisson regression model). The model assumes that the response variable Y has a Poisson distribution, and assumes that the

logarithm of its expected value are modeled by a linear combination of unknown parameters. Annual percent changes in rates are estimated by means of a linear regression of the logarithm of the respective rates on the mid-point of the calendar time period, weighted by the number of cases.

Jointpoint regression model is another way to characterize trends in cancer rates over time. This model uses statistical criteria to determine when and how often the annual percent change in rates (Hudson 1966; Lerman 1980). For cancer rates, it is fitted using joined log-linear segments, so each segment can be characterized using an APC. Cancer rates may rise gradually for a period of several years, rise sharply for several years after that, then drop gradually for the next several years. Finding the joinpoint model that best fits the data allows us to determine how long the APC remained constant, and when it changed. Statistical software called Joinpoint is available for the analysis of trends using joinpoint models, that is, models where several different lines are connected together at the "joinpoints". The user supplies the minimum and maximum number of joinpoints. The program starts with the minimum number of joinpoint and tests whether more joinpoints are statistically significant and must be added to the model. This enables the user to test whether an apparent change in trend is statistically significant (Kim et al., 2000).

Average Annual Percent Change (AAPC) is a summary measure of the trend over a pre-specified fixed interval. It allows us to use a single number to describe the average APCs over a period of multiple years. It is valid even if the joinpoint model indicates that there were changes in trends during those years. It is computed as a weighted average of the APC's from the joinpoint model, with the weights equal

to the length of the APC interval. AAPC is derived by estimating the underlying joinpoint model that best fits the data.

3. Summary of the present work

In the present thesis, different approaches such as i) without considering social value choices such as discount rate and age-weighting (i.e. zero discounting rate and uniform age-weighting) ii) only considering non-uniform discounting (at a rate of 3%) and uniform age-weights and iii) considering both non-zero discounting (at a rate of 3%) and non-uniform age-weights are used for estimating YLL due to female breast, cervix, ovary and corpus uteri cancers in Thiruvananthapuram. Further, YLD and DALY due to the above four cancers in Thiruvananthapuram are estimated with and without considering competing risk and based on different social value choices. Burden of these four cancers in Kerala till 2026 are estimated under certain assumptions and using the Thiruvananthapuram data. For estimating the future burden of breast and ovarian cancers, non-zero age-weighting and non-zero discounting values are finally chosen as the average age at diagnosis of these cancers are generally 5-10 years lower than cancers of the cervix and corpus uteri. Non-zero discounting values as well as all-cause mortality are chosen for all the four cancer sites.

Probability of developing the above four cancers are estimated using age-conditional probability (using competing risk model) and cumulative risk approach (without considering competing risk). Estimated annual percent change in incidence rates (rates that change at a constant percentage every year linearly on a log scale) and average annual percent change via joint point regression model for the

above four cancers are compared and these estimates are used for projection of cancer.

The thesis is organized into 12 chapters. Chapter 1 gives an introduction of the work with specific objectives. In chapter 2, data sources employed for estimating the various methods are provided. The data are incidence and mortality rates of female breast, cervix uteri, ovary and corpus uteri cancers reported in Thiruvananthapuram cancer registry from the year 1991 to 2010. Based on the census population for the years 1991 and 2001, the data for the years 2006, 2011, 2016, 2021 and 2026 are estimated using difference distribution method. In chapter 3, the various methods to estimate age-specific (ASp) and age-standardized (ASR) rates of cancer are provided. The advantages and limitations of each method are discussed. Incidence and mortality rates of female breast and reproductive tract cancers published in the literature are provided. These cancers have a high incidence (50-60% of all women cancers) in Thiruvananthapuram and mostly report in late stages for treatment and thus exercise an adverse influence on the productive role of women in the society. Factors affecting for delayed and late-stage reporting for breast and cervix cancers are published (Ali, **Mathew**, Rajan 2008; Kaku, **Mathew**, Rajan 2008; Paul, George, **Mathew**, 2010). Various risk factors associated with the development of breast cancer in Kerala are also published (**Mathew** et al., 2008; **Mathew** et al., 2009; Dey, ..., **Mathew**, 2009; Gajalakshmi, **Mathew**, ..., 2009).

In chapter 4, the methods to assess the burden of cancer in terms of DALY are provided and related studies on female breast and reproductive tract cancers published in the literature are provided. In chapter 5, the methods such as age-conditional probabilities and cumulative risk for estimating the probability of developing cancer and

related studies on female breast and reproductive tract cancers published in the literature are provided. The advantages and limitations of each method are discussed. In chapter 6, various analytical approaches used for estimating time trends in incidence and mortality rates of cancer are provided along with advantages and limitations of each method. Related studies on female breast and reproductive tract cancers published in the literature are provided. In chapter 7, various approaches employed for the projection of disease burden are provided along with advantages and disadvantages of methods.

In chapter 8, DALY for female breast and reproductive tract cancers in Thiruvananthapuram are compared by assuming with and without competing risk approach. Considering social value choices such as i) without discounting or uniform age weights ii) with non-zero discounting and uniform age weights and iii) non-zero discounting and non-uniform age weighting. YLL under the above three assumptions are 1031.96, 678.82, 555.29 for female breast cancer, 449.04, 295.32, 241.92 for cervix cancer, 270.04,186.65,142.03 for ovarian cancer and 213.4, 149.90, 111.14 for corpus uteri cancer per 10^5 women-years respectively. Non-zero discount rate of 3%, age-weighting correction constant (GBD standard value) of 0.1658, and age-weighting function (GBD standard value) of 4% are chosen (Gold et al., 1996; Murray and Acharya 1997).

With and without considering competing risk model, YLD was 123.47 and 133.22 for female breast cancer, 25.21 and 49.10 for cervix cancer, 22.96 and 25.68 for ovarian cancer and 35.97 and 19.87 for corpus uteri cancer per 10^5 women-years respectively. The YLD values obtained in the present thesis are slightly higher when not considered competing risk model for breast, cervix and ovarian

cancers. With and without considering competing risk model, DALY was 678.77 and 580.97 for female breast cancer, 267.14 and 291.03 for cervix cancer, 165.00 and 161.91 for ovarian cancer and 147.11 and 131.02 for corpus uteri cancer per 10^5 women-years respectively. DALY with and without considering competing risk model are not differed much and thus it was observed that in the absence of all cause-mortality data in cancer registries, burden of cancer can be estimated using incidence, mortality and case-fatality rate.

In chapter 9, the probability of developing the above cancers in Thiruvananthapuram by using age-conditional probability and cumulative risk approach are provided. Using age-conditional probability, the estimated probability of developing breast cancer was 2.61% and using cumulative risk approach, it was 2.89% during 0-64 years. The conditional-probability of dying due to breast cancer was 0.32% and using cumulative risk approach, it was 0.34% during 0-64 years. The conditional probability of developing cervix cancer was 0.69% and using cumulative risk approach, it was 0.79% during 0-64 years. The conditional-probability of dying due to cervix cancer was 0.16% and using cumulative risk approach, it was 0.18% during 0-64 years. The conditional probability of developing ovarian cancer was 0.58% and using cumulative risk approach, it was 0.68% during 0-64 years. The probability of dying due to ovarian cancer was 0.11% and using cumulative risk approach, it was 0.13% during 0-64 years. The conditional probability of developing corpus uteri cancer was 0.35% and using cumulative risk approach, it was 0.40% during 0-64 years. The conditional probability of dying due to corpus uteri cancer was 0.11% and using cumulative risk approach, it was 0.13% during 0-64 years.

Differences between the probabilities estimated according to the two methods were less than 10% for all the above cancer sites and in all the age-intervals. Slightly higher probability values were obtained when estimated using cumulative risk approach as this method is free from the assumption of competing causes of death. The interval 0-64 years is chosen arbitrarily to illustrate the difference in values between the two methods. It was observed that in the absence of all-cause mortality data in cancer registries, probability for developing or dying due to cancer can be estimated using cumulative risk approach.

In chapter 10, the annual percent changes in incidence of the above four cancers in Thiruvananthapuram were estimated by assuming linear and linear with join point changes in rates. Based on the incidence rates during the 20-year (1991-2010) period, it was observed that the estimated annual percent change (EAPC) was 3.8% ($p<0.05$) (assuming linearity). However, from the graphs, it was observed that the rates were not increased monotonically and thus strictly an assumption of linear increase may be biased. Considering two joinpoints, the average annual percent change (AAPC) for breast cancer incidence rate was 1.3% ($p<0.05$). Similarly the EAPC and AAPC for cervix cancer were -2.2% and -2.2%, for ovarian cancer were, 3.3% and 2.9% and corpus uteri cancer were 6.7% and 2.9% respectively.

Reporting APCs for each joinpoint segment provides a complete characterization of the trend over time. Prior to the development of the joinpoint and AAPC methodology, to characterize a trend over a fixed interval, a single regression line (on a log scale) over the fixed interval was fit, and the slope coefficient was then transformed to an APC. EAPC has two disadvantages over the AAPC. First, EAPC assumes linearity of the trend (on a log scale) over the interval, while the AAPC does not. Secondly, the AAPC can be used to characterize a short segment based

on a joinpoint model fit over a much longer series. This is especially advantageous for situations when the data are rare or data from a small geographic area. Trends in incidence rates of these cancers in India are published (Murthy,...,**Mathew** 2009; **Mathew** and George 2009; Murthy, ..., **Mathew**, 2011).

Chapter 11 provides the projected burden of the above four cancers in Kerala upto 2026. DALY due to female breast and reproductive tract cancers by five-year age groups for the period 2011, 2016 and 2021 and 2026 was estimated. The projection of these cancers have been done with the assumptions that (i) the age-specific incidence and mortality rate due to the above cancers as well as the general mortality rate in Thiruvananthapuram would be same in other areas in the state of Kerala and (ii) the estimated average annual percent change (AAPC) in Thiruvananthapuram for the above cancer sites would be the same in other areas in Kerala.

Chapter 12 provides summary and conclusion of the thesis. The thesis deals with a number of findings such as estimation of burden of female breast and reproductive tract cancers, probability of developing or dying due to these cancers, time trends in incidence rates of these cancers in Thiruvananthappuram and projection of burden of these cancers in Kerala. The results of this thesis are published in 11 research papers in international journals which are listed in the references. A few papers are communicated for publication.