
Modeling Age-specific Cancer Incidences Using Logistic Growth Equations: Implications for Data Collection

Xing-Rong Shen1*, Rui Feng2, Jing Chai1, Jing Cheng1, De-Bin Wang1,3

Abstract

Large scale secular registry or surveillance systems have been accumulating vast data that allow mathematical modeling of cancer incidence and mortality rates. Most contemporary models in this regard use time series and APC (age-period-cohort) methods and focus primarily on predicting or analyzing cancer epidemiology with little attention being paid to implications for designing cancer registry, surveillance or evaluation initiatives. This research models age-specific cancer incidence rates using logistic growth equations and explores their performance under different scenarios of data completeness in the hope of deriving clues for reshaping relevant data collection. The study used China Cancer Registry Report 2012 as the data source. It employed 3-parameter logistic growth equations and modeled the age-specific incidence rates of all and the top 10 cancers presented in the registry report. The study performed 3 types of modeling, namely full age-span by fitting, multiple 5-year-segment fitting and single-segment fitting. Measurement of model performance adopted adjusted goodness of fit that combines sum of squared residuals and relative errors. Both model simulation and performance evaluation utilized self-developed algorithms programed using C# language and MS Visual Studio 2008. For models built upon full age-span data, predicted age-specific cancer incidence rates fitted very well with observed values for most (except cervical and breast) cancers with estimated goodness of fit (Rs) being over 0.96. When a given cancer is concerned, the R values of the logistic growth model derived using observed data from urban residents was greater than or at least equal to that of the same model built on data from rural people. For models based on multiple-5-year-segment data, the Rs remained fairly high (over 0.89) until 3-fourths of the data segments were excluded. For models using a fixed length single-segment of observed data, the older the age covered by the corresponding data segment, the higher the resulting Rs. Logistic growth models describe age-specific incidence rates perfectly for most cancers and may be used to inform data collection for purposes of monitoring and analyzing cancer epidemic. Helped by appropriate logistic growth equations, the work volume of contemporary data collection, e.g., cancer registry and surveillance systems, may be reduced substantially.

Keywords: Cancer - incidence - models - logistic growth equations - data collection - China cancer registry

Asian J Cancer Prev, 15 (22), 9731-9737

Introduction

Cancers have long been a major cause of human death and diseases-adjusted life year loss (Ma et al., 2006). In 2002, new cancer cases and deaths accounted for 10.86 and 6.73 million respectively worldwide (Parkin et al., 2000). In 2005, over 7.6 million people died to cancer accounting for 13% of total death (Parkin et al., 2005). Predicted cases and deaths will rise to 15 and 10 million by 2020 (Parkin et al., 2001). WHO data showed that malignant tumors worldwide took up 5% of total burden caused by all diseases in 2005 (World Health Organization, 2006). More recent investigations revealed that cancers were the first death cause in cities and higher than cerebrovascular diseases and cardiopathy (China Ministry of Health, 2008). Estimated direct and indirect economic loss due to the disease was 11.323 and 6.000 billion USD respectively representing 4.67% of total medical cost (Wei, 2009).

Escalating cancer threats and harms have attracted tremendous efforts exploring the epidemiology of the diseases via large scale secular registry or surveillance systems (Goss et al., 2014; Ullrich et al., 2014). These efforts have been accumulating vast data that allow for establishment of mathematic models simulating cancer incidence and mortality rates for different age groups, time periods or cohorts. Leung and colleagues proposed a model for analyzing cervical cancer incidence using maximum likelihood and Bayesian methods and data from the Hong Kong Cancer Registry (Leung et al., 2006). Tyson and coworkers established a model incorporating the effects of age, year of diagnosis, and year of birth on incidence trends of renal cell carcinoma using data from
United States National Cancer Institute’s Surveillance, Epidemiology, and End Results public-use registry (Tyson et al., 2013). The most commonly adopted approaches for modeling cancer rates are time series and APC models (Meira et al., 2013; Ocana-Riola et al., 2013; Wang et al., 2014). Usually, time series models assume a Poisson distribution of cancer counts and include autoregressive error terms and/or time trends (Wingo et al., 1998; Knorr-Held et al., 2001); while APC models generally consists of three components, i.e., age (A), period (P), and cohort (C) (Jurgens et al., 2014). Of all the variables studied so far, age seems to have the highest effect on cancer mortality and incidence rates (Dyzmann-Sroka et al., 2014). So the performance of models depends heavily on how the influence of age on cancers is incorporated. With contemporary models, methods used simulating the relationship between age and cancer rates include mainly linear (Lee et al., 2011), polynomial (Wingo et al., 1998), piecewise linear (Kim et al., 2000), spline, log linear (Du et al., 2014) or power curves (Moller, 2004). Typical S-shaped line graphs of age-specific cancer incidence and mortality rates are clearly observable with almost all cancers worldwide (Bouchbika et al., 2013; Al-Hashimi et al., 2014; Wei et al., 2014). If S-line represents true general pattern, most methods (including linear, log linear, power curves) used in previous models may not fit well at least for some age ranges (e.g., under 35 or over 75 years). Some of the curves (e.g., piecewise linear and spline curves) may adequately approach any S-line. Yet this requires much more detailed data about observed age-specific cancer counts. Besides, most previous work in this regard focuses mainly on predicting or analyzing cancer epidemic with little attention being paid to informing relevant data collection.

This paper models age-specific cancer incidence rates using logistic growth equations. Although there are evidences that such equations describe well cancer cell proliferation under various conditions (Fory’s et al., 2003), publications linking them with age-specific cancer rates are limited. In particular, the paper explores the performance of logistic growth models under different scenarios of data completeness. This may reveal clues for reshaping contemporary data collection, e.g., cancer registry, surveillance or evaluation initiatives. Given the huge amount of scarce resources invested annually on these initiatives (Hutchison et al., 1997), they merit continuous scrutinize and refinement.

Materials and Methods

Data source

All source data used in this study came from China Cancer Registry Report 2012, the latest available annual report of the kind by far (He et al., 2012). It draws from data collected in 2009 by 72 sites throughout China covering 85.47 million urban and rural Chinese residents and provides incidence and mortality rates of all and over 20 specific cancers by age, gender and registry sties. A sample datasets was given in our previous paper (Chen et al., 2014) and detailed characteristics of the data will be described seperately.

Formulae used

Based on empirical observations of patterns with the reported cancer incidence rates along different ages for all and specific cancers, the study adopted a three parameter logistic growth equation (Formula 1). In this formula, \( t \) stands for age; and \( p_s \), cancer incidence rate for a given age \( t \); \( p_{st} \), the highest cancer incidence rate for all ages; \( k \), growth rate; while \( b \) serves as a baseline growth rate that determines the location of “the rapidly growing phase” of a S-curve along the age spectrum. In addition, the study used Formula 2 in identifying the most optimal model from a set of potential models for a given type of cancer and in evaluating the performance of the models selected. In Formula 2, \( R \) represents goodness of fit of the model under concern; while \( p_{st} \) and \( p_s \) stand for observed and simulated (or predicted) cancer incidence rate for a given age \( t \) respectively.

\[
P_t = \frac{P_{\text{max}} e^{pt}}{1 + e^{pt}}
\]  

(1)

\[
R = 1 - \sqrt{\frac{\sum (\text{Obs} - \text{Pred})^2}{\sum \text{Pred}^2}}
\]  

(2)

Selection of input data

In terms of how input data were selected, the study performed 3 types of modeling, namely full age-span fitting, multiple 5-year-segment fitting and single-segment fitting. Full age-span fitting utilized the registered cancer incidence data covering all the ages (i.e., from age 0 through to age 85); while the other two types of fitting, only part of the data. The multiple 5-year-segment fitting divided the whole age-span into segments consisting of 5 consecutive ages (e.g., ages 0-4, ages 5-9 etc.) first and then enter the corresponding observed age-specific cancer incidence rates for every other (e.g., ages 0-4, ages 10-14, ..., ages 70-74, ages 80-84), every other 2, every other 3, every other 4 and every other 5 segments into modeling respectively. With regard to single-segment fitting, it selected only one segment of ages and entered corresponding incidence rates into modeling. These segments covered 15, 30, 45, 60 and 75 consecutive ages respectively. Considering that location along the age-span covered by a same length of data segment may result in different model parameters, the study set beginning, middle and end as 3 criteria for selecting single segment data. For example, for the segment consisting of 15 ages, the study established 3 different logistic growth equations based on registered cancer incidence rates for ages 0-14 (beginning segment), ages 38-52 (middle segment) and ages 70-84 (end segment) respectively.

Algorithms for model building

Given that available statistical software does not allow for logistic growth modeling using segmental input data. The study employed a self-developed mini-program to perform all the computation. Written in C# language, the program runs on a webpage built with Microsoft Visual Studio 2008. For each model building, the webpage accepts an intended age set (e.g., \( \{61; 62; 63; 64; 65\} \))
and a corresponding set of observed cancer incidence rates (e.g., \((351; 362; 373; 384; 395)\), in \(1/100000\)) as input and then produces a best-fit parameter set (e.g., \(p = 0.08, b = 10.5, k = 0.17\)) and a goodness of fit (e.g., \(R = 0.98\)). This computation proceeds in 5 steps. Step 1 assumes a proper value range for each of the 3 parameters included in Formula 1 (i.e., \(0 ≤ p ≤ 0.8, 0 ≤ b ≤ 0.8, 0 ≤ k ≤ 0.8\)). Step 2 sets a small enough incremental value for each of the 3 parameters, i.e., \(1 \text{ for } p, \text{ max } = 200, b = 0.1, k = 0.01\) for \(b\) and \(0.01\) for \(k\), respectively and divides the ranges of \(p, b\) and \(k\) into 3 serial parameter sets, i.e., \(\{0, 1, 2, ..., 1 \times i, ..., 4999, 5000\}\), \(\{0, 0.1, 0.2, 0.3, ..., 0.1 \times j, ..., 19.8, 19.9, 20.0\}\) and \(\{0, 0.01, 0.02, ..., 0.01 \times n, ..., 0.78, 0.79, 0.80\}\); here \(i = 1, 2, 3, ..., 5000\) (i.e., \(5000/1\)), \(j = 1, 2, 3, ..., 200\) (i.e., \(20/0.1\)) and \(n = 1, 2, 3, ..., 80\) (i.e., \(0.8/0.01\)). Step 3 selects one element from each of the 3 serial parameter sets and generates a complete set (5000 \(\times 200 \times 80\) elements in total) of potential parameter combinations, i.e., \(\{p_{\text{max}} = 1, b = 0, k = 0\}\), \(\{p_{\text{max}} = 0, b = 0, k = 0.01\}\), ..., \(\{p_{\text{max}} = 5000, b = 20, k = 0.79\}\), \(\{p_{\text{max}} = 5000, b = 20, k = 0.80\}\). Step 4 uses Formula 2 and compares the goodness of fit between the registered cancer incidence set entered via the webpage and that predicted by Formula 1 using each of the potential parameter combinations. Step 5 outputs the parameter combination that has the largest \(R\).

**Results**

**Full age-span fitting models**

As shown in Table 1 and Figure 1, the majority of curves representing observed age-specific cancer incidence rates fit very well with predictions by logistic growth models with estimated goodness of fit (\(R\)) being over 0.96. Yet the \(R\) values for some types of cancer, e.g., cervical cancer and breast cancer, were quite low, ranging from 0.25 to 0.92. The 3 parameters defining the logistic growth models showed substantial variations. \(P_{\text{max}}\) ranged from 8 to 2248; \(b\), from 5.50 to 14.80; and \(k\), from 0.08 to 0.44.

**Multiple-5-year-segment fitting models**

Table 2 and Figure 2 displays findings from multiple-5-year-segment modeling. Goodness of fit (\(R\)) decreased as the number of data segments being left out increased. Yet, it remained fairly high even when only one fourth of segments of observed data were entered into modeling. And this phenomenon applied to all types of cancers. The differences between the \(R\) values of models for different cancer types (excluding cervical and breast cancer) built upon “every other segment” of observed data (Table 2, column 5) and that built upon “every other 4 segments” of

![Figure 1. Predicted vs Registered Age-specific Cancer Incidence Rates](image1)

**Table 1. Parameters and Goodness of Fit of Logistic Growth Models Based on Full-age-span Cancer Incidence Data**

<table>
<thead>
<tr>
<th>Type of cancer</th>
<th>Total</th>
<th>Males</th>
<th>Females</th>
<th>Urban</th>
<th>Rural</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(P_{\text{max}} )</td>
<td>(b)</td>
<td>(k)</td>
<td>(R)</td>
<td>(P_{\text{max}} )</td>
</tr>
<tr>
<td>Lung cancer</td>
<td>532</td>
<td>8.7</td>
<td>0.12</td>
<td>0.97</td>
<td>779</td>
</tr>
<tr>
<td>Esophageus cancer</td>
<td>129</td>
<td>10.6</td>
<td>0.17</td>
<td>0.98</td>
<td>200</td>
</tr>
<tr>
<td>Gastric cancer</td>
<td>245</td>
<td>8.6</td>
<td>0.13</td>
<td>0.97</td>
<td>338</td>
</tr>
<tr>
<td>Liver cancer</td>
<td>151</td>
<td>5.8</td>
<td>0.09</td>
<td>0.94</td>
<td>194</td>
</tr>
<tr>
<td>Colorectal cancer</td>
<td>272</td>
<td>8</td>
<td>0.11</td>
<td>0.96</td>
<td>333</td>
</tr>
<tr>
<td>Bladder cancer</td>
<td>91</td>
<td>9.3</td>
<td>0.12</td>
<td>0.96</td>
<td>160</td>
</tr>
<tr>
<td>Cervical cancer</td>
<td>10</td>
<td>13.5</td>
<td>0.39</td>
<td>0.73</td>
<td>NA</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>44</td>
<td>11.6</td>
<td>0.28</td>
<td>0.86</td>
<td>3</td>
</tr>
<tr>
<td>Pancreatic cancer</td>
<td>95</td>
<td>8.5</td>
<td>0.11</td>
<td>0.98</td>
<td>112</td>
</tr>
<tr>
<td>Nasopharyngeal cancer</td>
<td>8</td>
<td>6.8</td>
<td>0.16</td>
<td>0.92</td>
<td>12</td>
</tr>
<tr>
<td>All cancers</td>
<td>2064</td>
<td>7</td>
<td>0.1</td>
<td>0.96</td>
<td>2170</td>
</tr>
</tbody>
</table>

*Note: Source data came from age-specific incidence rates of top ten and all cancers from China cancer registry report 2012; \(P_{\text{max}}\), \(b\) and \(k\) represents the parameters in the logistic equation, \(y = P_{\text{max}}/(1+e^{-\text{kt}})\), where \(t\) stands for age and \(y\), incidence rate for age \(t\); \(R\) stands for goodness of fit between predicted and observed age-specific cancer incidence rates; NA stands for not applicable*
observed data (Table 2, column 17) ranged from only 0 to 0.05. However, starting from the column of “every 5 other segments”, the R values reduced dramatically. Similarly, although all the 3 parameters of the simulated logistic growth equations varied as the number of segments of data entered for modeling changed, most of these variations remained to a minimum extent (less than 10%) until the column of “every 4 other segments” and did not show clear decreasing or increasing trend.

Table 3. Parameters and goodness of fit of Logistic Growth Models Based on Single 5-age-segment Cancer Incidence Data

<table>
<thead>
<tr>
<th>Type of cancer</th>
<th>Every other segment</th>
<th>Every other 2 segments</th>
<th>Every other 3 segments</th>
<th>Every other 4 segment</th>
<th>Every other 5 segments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \beta )</td>
<td>( k )</td>
<td>( R )</td>
<td>( \beta )</td>
<td>( k )</td>
</tr>
<tr>
<td>Lung cancer</td>
<td>485 9.2 0.13 0.96</td>
<td>517 9.3 0.13 0.96</td>
<td>484 9.2 0.13 0.96</td>
<td>530 8.6 0.12 0.95</td>
<td>224 10.4 0.17 0.63</td>
</tr>
<tr>
<td>Esophagus cancer</td>
<td>132 10.1 0.16 0.98</td>
<td>126 11.1 0.18 0.98</td>
<td>130 10.8 0.19 0.96</td>
<td>127 10.5 0.17 0.96</td>
<td>105 14.4 0.24 0.88</td>
</tr>
<tr>
<td>Gastric cancer</td>
<td>348 8.5 0.13 0.97</td>
<td>250 8.6 0.13 0.97</td>
<td>233 9.1 0.14 0.97</td>
<td>251 8.5 0.13 0.96</td>
<td>178 9.2 0.15 0.84</td>
</tr>
<tr>
<td>Liver cancer</td>
<td>167 5.4 0.08 0.94</td>
<td>156 5.9 0.09 0.94</td>
<td>158 5.9 0.09 0.94</td>
<td>129 7 0.12 0.89</td>
<td>90 7.6 0.14 0.78</td>
</tr>
<tr>
<td>Colorectal cancer</td>
<td>248 8.5 0.12 0.95</td>
<td>291 8.1 0.11 0.95</td>
<td>269 8 0.11 0.96</td>
<td>262 7.8 0.11 0.93</td>
<td>120 8.5 0.14 0.65</td>
</tr>
<tr>
<td>Bladder cancer</td>
<td>90 9.3 0.12 0.96</td>
<td>105 9.5 0.12 0.94</td>
<td>81 9.8 0.13 0.95</td>
<td>95 9.3 0.12 0.94</td>
<td>31 8.5 0.13 0.55</td>
</tr>
<tr>
<td>Cervical cancer</td>
<td>10 18.5 0.54 0.73</td>
<td>10 14.8 0.43 0.73</td>
<td>10 19.8 0.67 0.67</td>
<td>10 14.7 0.44 0.73</td>
<td>10 9.5 0.27 0.72</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>44 12.1 0.29 0.9</td>
<td>49 10.9 0.27 0.9</td>
<td>42 11.7 0.28 0.89</td>
<td>44 16.3 0.47 0.74</td>
<td>46 9.2 0.22 0.89</td>
</tr>
<tr>
<td>Pancreatic cancer</td>
<td>95 8.5 0.11 0.97</td>
<td>89 8.9 0.12 0.97</td>
<td>83 9 0.12 0.97</td>
<td>90 8.4 0.11 0.97</td>
<td>33 10.7 0.17 0.62</td>
</tr>
<tr>
<td>Nasopharyngeal cancer</td>
<td>8 6.8 0.16 0.92</td>
<td>8 7.2 0.17 0.92</td>
<td>8 8.5 0.2 0.91</td>
<td>7 8.4 0.21 0.89</td>
<td>9 6.3 0.14 0.87</td>
</tr>
<tr>
<td>All cancers</td>
<td>2248 6.5 0.09 0.97</td>
<td>2180 7.1 0.1 0.96</td>
<td>2229 6.6 0.09 0.97</td>
<td>1989 6.8 0.1 0.94</td>
<td>1151 7.2 0.12 0.75</td>
</tr>
</tbody>
</table>

Note: Source data came from age-specific incidence rates of top ten and all cancers from China cancer registry report 2012; \( \beta \) and \( k \) represents the parameters in the logistic equation, \( y = \frac{P_{\text{max}}}{1 + e^{-\beta t + k}} \), where \( t \) stands for age and \( y \), incidence rate for age \( t \); \( R \) stands for goodness of fit between predicted and observed age-specific cancer incidence rates.

Single-segment fitting models

Table 3 and Figure 3 resulted from single-segment fittings. Goodness of fit (R) increased as the length of the data segment increased and this increase was dependent on the location of the segment of data entered for fitting. For a same length of segment (e.g., 15 ages), the older the age covered by the corresponding data segment, the higher the resulting \( R \). As for the segment covering the oldest part of age-span, all the \( R \) values turned out to be very high. The
modeled parameters were also linked to the length and age-range covered by the data segment. For data segment covering beginning ages, all the 3 parameters increased as the length changed from 15 ages to 75 ages; while for data segment covering the end ages, $P_{\text{max}}$ increased yet $b$ and $k$ decreased as the length increased.

**Discussion**

Although typical S-shaped line graphs of age-specific cancer incidence rates are clearly observable with almost all cancer registry and other relevant epidemiological reports worldwide, their relations with logistic growth equations have not been fully addressed. The current study demonstrated that logistic growth models perfectly describe the incidence rates along different age groups for most type of cancers. This may be explained by: a) onset of clinically detectable cancers results from the counteraction between cancer cell occurrence and removal (Baker et al., 2013); b) cancer cell occurs after a normal somatic cell has experienced multiple times (say n times) of damages due to exposure to same or different risk factors (Shaukat et al., 2013); c) a certain level of risk exposure defines a corresponding chance ($q$) for a normal somatic cell to get one time damage and hence the chance ($q^p$) for an innate cell to mutate into cancer cell in an unit time period; d) given c, as time ($t$) passes by and somatic cell gets damaged for more and more times, its chance ($p$) for becoming malignant increases exponentially ($p = q^n$); e) level of life spectrum exposure to cancer risk factors starts relatively low at birth, increases during childhood and adolescence (due initiation of unhealthy or unprotected behaviors), remains the highest in adulthood and begins to decrease gradually in late lifetime (due to reduced smoking, drinking etc.) (Katulanda et al., 2014; Chockalingam et al., 2013); f) cancer cell removal or immunity manifests similar lifetime trend as risk exposure (Wu et al., 2012). Therefore, the early low and relatively stable phase of the S-shaped age-specific cancer rates may reflect the combined effect of low cancer cell occurrence vs. high immunity; while the rapidly growing part, exponentially increasing occurrence vs. high and stable immunity; and the late high and relatively stable stage, diminishing occurrence due to reduced risk exposure vs. downward immunity.

Linking logistic growth law with age-specific cancer rates leads to a plausible thinking that description of cancer incidence or mortality rates along the whole age span is to estimate the parameters involved in the equations rather than uncover counts for each of the ages. Such a shift of focus may result in great resource reduction, since logistic equations generally involve only a few parameters (e.g., 3 parameters in our cases) and estimation of these requires much less data than what have usually been collected. This is of particular significance to cancer registry. As suggested by our simulations (Table 2 and Figure 2), the work volume of current China national cancer registry could be reduced by 3 fourths without severely damage its capacity in producing age-specific cancer incidence rates. This should also apply to other registries. Given that over fifty countries have large scale operating cancer registry systems that consume huge amount of scarce resources year by year (Izquierdo et al., 2000; Tangka et al., 2010), a growth model-guided rethinking merits special attention. Even though segmental cancer registry may sound unacceptable to some, the findings suggest priority age groups for monitoring and controlling data quality of registry systems.

Logistic growth analysis may also inform data collection for intervention or hypothesis assessments. As shown in Table 3 and Figure 3, for a same length of data segment, the older the age covered by the data, the higher the goodness of fit of the resulting model. This suggests that, for studies evaluating the effect of an intervention or an influencing factor on cancer rates using limited age groups, backward sampling (i.e., start to choose from the oldest age group backward to younger ones) may work better than forward selection (from age 0-5 to 6-10 and then to 11-15 etc.). For studies that have yielded data showing differences in cancer rates between two groups (say, intervention vs control) of middle ages (say, ages 30-59), simulated logistic growth equations may be used to measure extended difference (say for ages 60-69, or even 60 and over) between the two groups. However, the goodness of fit of models based on data covering middle segment of ages is only moderate.

In addition, logistic growth equations may help assessing data collections biases and/or errors under certain circumstances. If there are sufficient evidences to believe that certain age-specific cancer rates follow logistic growth law, then the goodness of fit estimations (Rs) can also be viewed as a quality indicator of the observed cancer counts. Of the ten cancers included in Table 1, cervical and breast cancers showed clear deviations from logistic equations. By excluding these two cancers, all the cancer-specific pairs of Rs (Table 1, column 17 vs 21) showed a consistent trend, i.e., for any given cancer, the R of the model built upon observed data from urban residences was higher than or at least equal to that from rural people. This may indicate better cancer registry in urban than in rural China. The Rs for models of different cancers witnessed much greater variations ranging from 0.92 for nasopharyngeal cancer to 0.98 for esophagus and pancreatic cancers (Table 1, column 5). This suggests a need for tailored data quality control or...
improvement with special attention being paid to cancers with the lowest Rs. The varied biases and errors in the rates for different cancers in our case may be attributed to a whole range of reasons including number of cases registered (e.g., too few for nasopharyngeal cancer), physical symptoms and sings, easiness to get cancer tissues for pathologic diagnosis, availability of auxiliary examination techniques etc.

Finally, readers are cautioned about a number of issues. First, this study used only simplest logistic growth equations and they do not fit very well with the observed data for some cancers, e.g., cervical and breast cancers. Such problems can be solved by adding more parameters and introducing more sophisticated growth equations. Second, parameters presented in this paper were all average estimates derived from pooled cancer counts reported by 72 CNCR sites in 2009. Age-specific incidence and mortality bands with means and 95% confidence intervals rather than single mean estimates may be produced by building similar set of logistic growth equations using the data from each CNCR sites (72 sets in total) and then performing bootstrap re-sampling and jackknife-correction (Dexter et al., 2013; Yu et al., 2013). Third, this paper focuses primarily on implications for data collection without any attention being paid to identifying trends and components with the cancer rates. Forth, apart from goodness of fit, this paper did not provide other performance indicators (sensitivity, specificity etc.) of the models used due to space limit. Most of these will be addressed separately in a forthcoming paper titled “modeling age-specific cancer incidence using logistic growth equations: jackknife-corrected bootstrap estimates”.

Acknowledgements

This paper was funded by the Natural Science Foundation of China (Grant Number 81172201). All authors declare no conflicts of interest.

References


Manag, 37, 152-5.