

Spatial and temporal evolution of liver cancer incidence

Yuliang Xie^{1, #}, Han Lin^{2, *, #}, Wenhui Gao^{3, #}

¹ School of Mathematics and Statistics, Shandong University, Weihai, China, 264209

² College of Mathematics and Computer Science, Shantou University, Shantou, China, 515063

³ School of Mathematics and Information Technology, Yuncheng university, Yuncheng, China, 044000

* Corresponding author: 21hlin3@stu.edu.cn

#These authors contributed equally.

Abstract. Liver cancer has become a global public health problem and is the fourth leading cause of death and the sixth most common cancer worldwide. The article comprehensively constructed a system of indicators to measure the level of liver cancer incidence from three aspects: economy, diet, and health indicators, and analyzed the influencing factors of liver cancer incidence in 204 countries and regions in the world and the geographical characteristics of the level by using multivariate statistical analyses such as the coefficient of variation method and geodetector. The study shows that (1) the incidence rate of liver cancer is lower in countries with higher GDP levels and urbanized population rates. (2) The combination of caloric intake and protein intake had the largest interaction effect on liver cancer incidence. (3) Excessive fat and protein intake will increase the burden on the liver and cause an increase in the incidence of liver cancer.

Keywords: Geodetector, Coefficient of variation method, Liver cancer.

1. Introduction

Hepatocellular carcinoma is a common malignant tumor, and studies on its prevalence worldwide have important background and significance. Liver cancer has emerged as a global public health problem and is the fourth leading cause of death and the sixth most common cancer worldwide. In 2018, liver cancer caused about 780,000 deaths. The global variation in the incidence of liver cancer is significant. Asian countries, particularly those in East and Southeast Asia, continue to have the highest incidence of liver cancer. While incidence rates have been declining in most of these high-risk countries in recent years, they have been increasing in India and several low-risk countries in Africa, Europe, the Americas, and Oceania.

Therefore, studying the incidence of liver cancer is essential for developing effective prevention and intervention measures to reduce the number of liver cancer cases. The incidence of liver cancer varies significantly across the globe. Asia, Africa and Central and South America have lower incidence rates, while North America and Europe have higher rates. Understanding these geographic differences can help identify possible risk factors and develop targeted preventive measures.

Wang Jinfeng, Xu Chengdong et al. mainly studied the principles of geoprobes, summarized their characteristics and applications, and facilitated other scholars to use geoprobes to recognize, explore and utilize spatial variability [1]. Sun Huangping, Huang Zhenfang, Xu Dongdong and others used the coupling coordination degree model, geodetector and other methods to study the spatial characteristics of the coupling coordination degree of urbanization and ecological environment of the Pan-Yangtze River Delta urban agglomeration from 2002 to 2014, and got the conclusion that the coupling degree shows an inverted "U"-shaped curve, and that there is significant spatial clustering of the coupling degree of coordination, and that the main influences of spatial differentiation are the value-added of secondary and tertiary industries, which accounted for the increase in value-added of secondary and tertiary industries. The main influence of spatial differentiation is the proportion of added value of secondary and tertiary industries in GDP [2]. Zou Lei, Liu Huiyuan, Wang Feiyu et al. analyzed the characteristics of regional differences in the level of green development of city clusters in the middle reaches of the Yangtze River and its influencing factors from 2008 to 2018 by

using the comprehensive indicator evaluation model, spatial autocorrelation analysis and geodetector [3]. Zhao Wei, Lin Jian, Wang Shufang and others considered 13 indicators of human activities in order to evaluate the impact of human activities on the groundwater environment in various subsystems of the groundwater in Beijing, and used the coefficient of variation method to determine the weights of the indicators for the comprehensive evaluation, and obtained the results that the impacts of human activities on the first aquifer group were greater than that of the second aquifer group, and that the impacts on the groundwater subsystems of the Yongding River and the Chaobai River were greater than that on other groundwater subsystems [4]. Jae-Jun concluded that liver cancer is one of the leading causes of cancer-related deaths in Korea and that liver cancer has a high incidence and mortality rate among young patients compared to other cancers, causing a considerable social burden, but the interest of researchers and health policy makers in liver cancer is low. [5] According to the current research status, there are fewer studies on the combination of diseases such as liver cancer with geo-detectors, in view of this, this paper will construct a comprehensive index evaluation system for liver cancer incidence from seven indicators, namely, urbanized population rate, GDP level, annual per capita calorie intake, annual per capita protein intake, annual per capita fat intake, annual per capita vegetable intake, and annual per capita BMI, using the Multivariate statistical methods such as the coefficient of variation method and geographic detector were used to conduct an in-depth study on the influencing factors of the incidence rate of hepatocellular carcinoma as well as the spatial-temporal evolution of the incidence rate of hepatocellular carcinoma in the period of 2002-2019, and the results of the study were intended to provide theoretical basis for the prevention of hepatocellular carcinoma as well as the reduction of its incidence rate[10].

2. Methodology

2.1. The idea of spatio-temporal evolution

In order to initially observe the statistical distribution of global liver cancer incidence rate, we use the method of spatio-temporal evolution to analyze the distribution trend of global liver cancer incidence rate. We first conducted data collection to collect data on liver cancer incidence rates in different regions and years. We obtained them from the World Health Organization (WHO), national cancer registry databases, health departments, and medical institutions, and cleaned the collected data to ensure the accuracy and consistency of the data. For missing values and outliers in the dataset, interpolation can be used to estimate the incidence of liver cancer in order to obtain complete data globally, making the data available for subsequent analyses for regions with missing data. Next, we visualized the data using methods such as graphs or maps in order to observe the temporal trends in liver cancer incidence. Regarding temporal trend analysis, we used time series methods to explore the changes in liver cancer incidence over time, during which time trends can be analyzed using methods such as linear regression and trend graphs.[5-9] For spatial analysis, we can use tools such as the GEODA package to correlate liver cancer incidence rate data with geographic locations to analyze the differences between different regions and spatial distribution patterns. Finally, we perform trend prediction and statistical analysis. Based on the results of the temporal trend and spatial analysis, we can try to predict the future trend of liver cancer incidence and use appropriate statistical methods, such as correlation analysis and spatial autocorrelation, to explore the relationship between liver cancer incidence and other possible influencing factors.

2.2. Principle of geodetector

Geo-detector is used to analyze spatial stratified heterogeneity, which mainly consists of four detectors (factor detector, interaction detector, risk zone detector, and ecological detector).Geo-detector is a spatial analysis method that detects spatial differentiation as well as reveals the driving forces behind it, and is widely used to conduct driver analysis and factor analysis. The core idea is based on the assumption that if an independent variable has an important effect on a dependent variable, then the spatial distributions of the independent and dependent variables should be similar.

Geographic divergence can be statistically analyzed using geodetectors, which have two major advantages: first, geodetectors can detect both numerical and qualitative data; second, they can detect two-factor interactions on the dependent variable. Geographic detector by calculating and comparing the q-value of each single factor and the q-value of the superposition of the two factors, we can determine whether there is an interaction between the two factors, as well as the strength of the interaction, the direction of the interaction, linear or nonlinear, and so on. The superposition of two factors includes both multiplicative and other relationships, as long as there is a relationship, it can be tested.

2.3. Data collection and pre-processing

For the collection of liver cancer incidence data, we obtained datasets from the World Health Organization (WHO), health departments of various countries and cancer registry databases medical journals and academic publications, and took the liver cancer incidence data from 2002 to 2019 for the study. For the acquired dataset, we performed data preprocessing. Data preprocessing is an important step to ensure data quality and reliability. In preprocessing the liver cancer incidence data, for the missing values in the data, we choose appropriate methods to fill the missing values or delete the data points with more missing values according to the situation. For outliers, the data are detected for outliers, the abnormal data that may affect the analysis results are identified, and the interpolation algorithm is used to replace him or delete him directly. If the data come from different regions or different years, there may be differences in the magnitude of the data, so data standardization is needed so that the data can be compared on the same scale.

3. Result

3.1. Trends in spatial and temporal evolution

In terms of temporal trends, the scope of this study covers 204 country regions. These include the world's five largest economies in 2022, namely the United States, China, Japan, Germany, and the United Kingdom, which accounted for 24.39%, 16.29%, 5.84%, 4.46%, and 3.26% of global GDP in 2022. The five most populous countries are China, India, the United States, Indonesia, and Pakistan. In 2021, the shares of these my five countries of the total population of the 204 countries and territories counted were 1,412 (17.9%), 1,407 million (17.8%), 370 million (4.2%), 270 million (3.55), 230 million (2.9%). 2008 , 2014, and global prevalence of all types of cancer in 2019, with spatial distribution as shown in Fig 1. Figure 1 is obtained from the global liver cancer data processed using the gis software

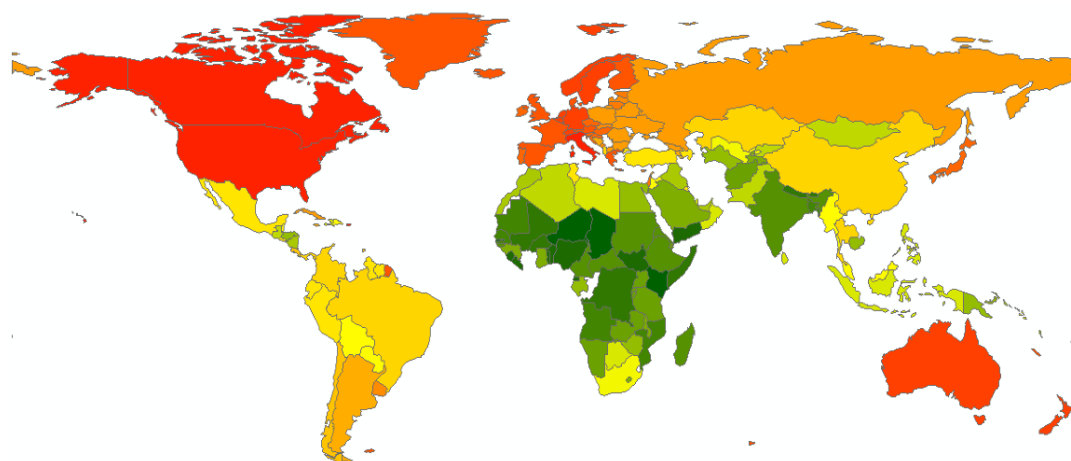


Figure 1. Global prevalence of various types of cancer

Overall, the global spatial distribution of the prevalence of liver cancer has been generally stable over the past 17 years, although national and regional variations are evident. Studies indicate that the

prevalence of liver cancer is lower in central and western Africa. Liver cancer prevalence is higher in Oceania, Eastern Asia, and North America and is on the rise. In 2019, the five countries with the highest global cancer prevalence were Mongolia (68.22), Republic of Korea (38.14), Japan (36.51), Thailand (35.41), and Tonga (19.21). The average liver cancer prevalence was higher in the European region, but the lowest coefficient of variation in the European region was 0.553, 0.521, 0.41, 0.41 in 2002, 2008, 2015, and 2019, respectively, which suggests that the vast majority of the European region has a serious liver cancer problem with very little variation among countries. Africa has the lowest liver cancer prevalence among all continents, but the coefficient of variation tends to stabilize 1.198, 1.178, 1.189, and 1.188 in 2002, 2008, 2015, and 2019, respectively, which suggests that the regional distributions of liver cancer prevalence among African countries tend to stabilize. For the observation of cancer prevalence in Asia, we found that although the prevalence of liver cancer is more serious in Asia, the coefficient of variation is the highest among all continents and shows a certain upward trend (1.075, 1.201, 1.739, and 1.265 in 2002, 2008, 2015, and 2019, respectively), which suggests that there is a huge difference in the prevalence of liver cancer among countries, and that this difference gradually increases over time. If the global average of liver cancer prevalence in countries and regions (5.71) is taken as the world average, the following findings are made. Based on the 2019 data, it is observed that the incidence of liver cancer in the East Asian region is very serious, with four of the five countries with the worst liver cancer in the world coming from the East Asian region. Liver cancer rates in the North American region are not encouraging either, mainly due to the high rates in Canada and the United States. The average cancer liver cancer prevalence rate of countries in Oceania is located in the first place in the world (7.77), which is mainly caused by the high incidence rates in Tonga and Australia (the cancer incidence rates in Australia in 2002, 2008, 2015, and 2019 were 4.41, 5.82, 7.57, and 7.32, respectively, which showed an upward trend, suggesting that the situation of liver cancer disease in Australia is increasingly serious).

3.2. Analysis of model construction

(1) Spatial heterogeneity:

In the process of exploring the spatial heterogeneity and interaction effects associated with the incidence of liver cancer, the data on the incidence of liver cancer in 2019 were utilized to carry out the study. Liver cancer incidence (abbreviated as Y) was classified into five categories: 0-3, 3-5, 5-7, 7-9, and 9-12, which are represented by the numerical labels 1 to 5, respectively. X1-X7 correspond to: urbanized population rate, GDP level, annual per capita calorie intake, annual per capita protein intake, annual per capita fat intake, annual per capita vegetable intake, and annual per capita BMI, respectively.

We performed a spatial heterogeneity analysis using the GD package in R. The calculated q-statistic for spatial heterogeneity of attribute variable Y was 0.4708775 and the corresponding statistical test p-value was highly significant, as shown in Table 1.

Table 1. Spatial heterogeneity results

	X1	X2	X3	X4	X5	X6	X7
q-statistic	0.016	0.018	0.224	0.287	0.361	0.139	0.044

(2) Spatial interactivity:

X1-X7 were all categorized into 5 categories based on the data distribution map, with 1-5 as numbering assignments.

Where X1-X7 correspond to: urbanized population rate, GDP level, annual per capita calorie intake, annual per capita protein intake, annual per capita fat intake, annual per capita vegetable intake, and annual per capita BMI, respectively.

From the economic aspect, as shown in Table 2, the urbanized population rate (X1) had a greater impact on liver cancer incidence than the GDP level (X2). The interaction effect of the two on liver cancer incidence is smaller than the univariate effect, i.e., countries with higher levels of both GDP and urbanized population will have lower liver cancer incidence.

Table 2. Interaction effects of X1, X2

	X1	X2
X1	0.0123	0.0990
X2	0.0990	0.0110

From the dietary aspect, as shown in Table 3, among calorie, protein, and fat intake, fat intake (X5) had the largest effect on liver cancer incidence, implying that high fat intake in daily life may increase the prevalence of liver cancer, and the combination of protein intake (X4) and calorie intake (X3) had the largest interaction effect on liver cancer incidence, suggesting that simultaneous intake of high protein and high calorie intake may increase liver cancer prevalence.

Table 3. Interaction effect of X3, X4, X5

	X3	X4	X5
X3	0.0419	0.2163	0.2054
X4	0.2163	0.0865	0.1461
X5	0.2054	0.1461	0.1002

(3) Ecological detection results:

As Fig 2 shows, in terms of urbanized population rate, there is a significant difference with other factors such as X3-X7: annual per capita calorie intake, annual per capita protein intake, annual per capita fat intake, annual per capita vegetable intake, and annual per capita BMI.

	X1	X2	X3	X4	X5	X6	X7
X1	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE
X2	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE
X3	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE
X4	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE
X5	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE
X6	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE
X7	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE

Figure 2. Analysis of ecological detection results

(4) Risk zone detection results:

As shown in Fig 3, there is no significant difference between the spreading degree of each type of sub-district in the factor of urbanized population rate, (N means there is no significant difference, and Y means there is a significant difference); as shown in Fig 4, there is only a significant difference between the spreading degree of the first sub-district and the third sub-district in the factor of GDP level.

X1					
	1	2	3	4	5
	6.032847	6.836896	5.559991	4.495933	5.8822
Sig. t test:	0.05				
	1	2	3	4	5
1					
2	N				
3	N	N			
4	N	N	N		
5	N	N	N	N	

Figure 3. Urbanized population rate factor

X2					
	1	2	3	4	5
	5.599569	7.205557	3.091637	5.24822	4.679533
Sig. t test:	0.05				
	1	2	3	4	5
1					
2	N				
3	Y	N			
4	N	N	N		
5	N	N	N	N	

Figure 4. GDP level factor and degree

As shown in Fig 5, there is no significant difference in the spread of the annual per capita calorie intake factor between the second and third sub-districts, the third and fourth sub-districts, and the fourth and fifth sub-districts. As shown in Fig 6, in the factor of annual per capita protein intake, there is a significant difference in the degree of spread between the first and the fourth and fifth divisions, and between the third and the fourth and fifth divisions.

X3				
1	2	3	4	5
1. 230657	2. 739256	6. 063654	6. 082425	8. 091921
Sig. t test: 0.05				
1	2	3	4	5
2 Y				
3 Y	N			
4 Y	Y	N		
5 Y	Y	N	N	

Figure 5. Per capita calorie intake

X4				
1	2	3	4	5
3. 59784	4. 425193	3. 361533	8. 090899	8. 059378
Sig. t test: 0.05				
1	2	3	4	5
2 N				
3 N	N			
4 Y	N	Y		
5 Y	N	Y	N	

Figure 6. Annual per capita protein intake

It can be known from Fig 7 that there is a significant difference in annual per capita fat intake between sub-division I and V, and between sub-division III and V in terms of degree of spread. Fig 8 shows that there is a significant difference in the spread of vegetable intake per capita per year between sub division I and sub divisions III, IV and V.

X5				
1	2	3	4	5
2. 801067	4. 653975	4. 576432	4. 837406	8. 756244
Sig. t test: 0.05				
1	2	3	4	5
2 N				
3 N	N			
4 N	N	N		
5 Y	N	Y	N	

Figure 7. Annual per capita fat intake

X6				
1	2	3	4	5
2. 473555	5. 929067	6. 469746	4. 955467	6. 015834
Sig. t test: 0.05				
1	2	3	4	5
2 N				
3 Y	N			
4 Y	N	N		
5 Y	N	N	N	

Figure 8. Annual per capita vegetable intake

Finally, analyzing the significance difference of annual per capita BMI against all spread degrees, it can be seen from Fig 9 that annual per capita BMI only (X7) is not significant against all spread degrees.

X7				
1	2	3	4	5
9. 803935	4. 113128	5. 788722	4. 727371	10. 82032
Sig. t test: 0.05				
1	2	3	4	5
2 N				
3 N	N			
4 N	N	N		
5 N	N	N	N	

Figure 9. Significance of per capita BMI value and spread in the year

3.3. Analysis of results

From the economic aspect, the effect of urbanized population rate on liver cancer incidence was more significant compared to GDP level. Further analysis showed that the interaction effect of urbanized population rate and GDP level on liver cancer incidence was small. In other words, liver cancer incidence was lower in countries with higher GDP levels and urbanized population rates.

In terms of diet, we found that fat intake had the most significant effect on liver cancer incidence, suggesting that high daily fat intake may increase the prevalence of liver cancer. In addition, the combination of calorie intake and protein intake had the largest interaction effect on liver cancer incidence, implying that simultaneous high protein and high calorie intake may increase the prevalence of liver cancer.

BMI can be used as an indicator of whether height and weight are standardized. After the interaction effect test, the largest interaction effect on liver cancer prevalence was found when both fat intake and BMI were higher, indicating that the prevalence of liver cancer would be higher in countries where both fat intake and BMI were higher, and the smallest interaction effect on liver cancer prevalence was found when both vegetable intake and BMI were higher, and it could be found that with the increase of vegetable It can be found that with the increase of vegetable intake, the effect on increasing the prevalence of liver cancer is small despite the simultaneous increase of BMI, and in order to reduce the risk of liver cancer, it can be encouraged to consume more vegetables and pay attention to the control of BMI.

The results of ecological detection showed that in terms of urbanized population rate, there were significant differences compared with other factors such as annual per capita calorie intake, annual per capita protein intake, annual per capita fat intake, annual per capita vegetable intake and annual per capita BMI. Among them, annual per capita fat intake was the most influential factor, followed by annual per capita protein intake, so excessive fat and protein intake would increase the burden on the liver and cause an increase in the incidence of liver cancer.

The results of risk zone detection showed that there was no significant difference between the spread of each type of partition in the urbanized population rate factor. For the GDP level factor, there is only a significant difference in the degree of spread between the first and third sub-districts.

4. Discussion

This paper mainly studies the overall incidence of liver cancer and its influencing factors, and lacks a more detailed analysis of particular countries or regions. From the results analyzed in this paper, the incidence rate of liver cancer is more serious in more economically developed regions, such as Europe, and thus it is speculated that liver cancer is a kind of rich and expensive disease. From the data, it appears that the incidence of liver cancer is lower in the African region, but due to the degree of difficulty as well as the complexity of data collection in the African region, the conclusions for the African region lack more solid support. From the data collected, the selected vegetable intake data was not broken down by type of vegetable, which could be further explored to see what type of vegetables consumed would be more helpful in suppressing the incidence of liver cancer. From the results of the analysis, the higher the BMI index, the higher the risk of liver cancer, but a deeper study on the BMI range of the lowest liver cancer incidence rate. From the direction of this paper, there is a lack of in-depth discussion on the possible "siphon effect", i.e., people from cities with high liver cancer incidence rates will move to cities with lower liver cancer incidence rates.

5. Conclusion

(1) As a whole, the spatial distribution of global liver cancer incidence is generally stable, and the incidence of liver cancer is on the rise. Locally, the difference in liver cancer incidence rates between countries in the European region is small, but the liver cancer rate is serious. In contrast, African countries have globally low liver cancer incidence rates and all African countries have stable low liver cancer rates. The above results are constructed in this paper to build a comprehensive evaluation index system construction, indicator data are mostly statistical data, lack of field research data, due to the inconsistency of the statistical caliber of each country, individual years are missing, to a certain extent, affecting the results of the analysis of the data.

(2) Using geographic detectors, it was found that the main influencing factors for the spatial differentiation of global liver cancer incidence rate were urbanized population rate, GDP level, annual per capita calorie intake, annual per capita protein intake, annual per capita fat intake, annual per capita vegetable intake, and annual per capita BMI, and all of them were characterized by obvious spatial and temporal heterogeneity. In conclusion, the incidence of hepatocellular carcinoma is the result of multiple factors interacting with economic factors, natural factors and dietary structure.

References

- [1] WANG Jinfeng, XU Chengdong. Geoprobes: Principles and prospects [J]. Journal of Geography, 2017, 72 (01): 116 - 134.
- [2] Sun Huangping, Huang Zhenfang, Xu Dongdong et al. Spatial characteristics and driving mechanism of urbanization and ecological environment coupling in pan-Yangtze Delta city cluster [J]. Economic Geography, 2017, 37 (02): 163 - 170+186.
- [3] Zou Lei; Liu Huiyuan; Wang Feiyu; Chen Ting; Dong Yi. Regional differences in green development level of city clusters in the middle reaches of the Yangtze River and their influencing factors [J]. Chinese Science: Earth Science, 2022, 52 (08): 1462 - 1475.
- [4] ZHAO Wei, LIN Jian, WANG Shufang et al. Evaluation of the impact of human activities on groundwater environment by coefficient of variation method [J]. Environmental Science, 2013, 34 (04): 1277 - 1283.
- [5] Shim J J. Current Status of Research on Liver Cancer in Korea [J]. Journal of Liver Cancer, 2017, 17 (2): 105 - 110. DOI: 10.17998/jlc.17. 2. 105.
- [6] Song, Y., Wang, J.F., Ge, Y., et al. An optimal parameters-based geographical detector model enhances geographic characteristics of explanatory variables for spatial heterogeneity analysis: cases with different types of spatial data. GIScience & Remote Sensing, 2020. 57 (5): 593 - 610.
- [7] Wang, J. F., Li, X. H., Christakos, G., Liao, Y. L., et al. Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun Region, China. International Journal of Geographical Information Science, 2010. 24 (1), 107 - 127.
- [8] Song, Y., Wu, P. An interactive detector for spatial associations. International Journal of Geographical Information Science, 2021.
- [9] Song, Y., Wu, P., Gilmore, D., et al. A Spatial Heterogeneity-Based Segmentation Model for Analyzing Road Deterioration Network Data in Multi-Scale Infrastructure Systems. IEEE Transactions on Intelligent Transportation Systems, 2020.
- [10] Petrick J L, Braunlin M, Laversanne M, et al. international trends in liver cancer incidence, overall and by histologic subtype, 1978 – 2007 [J]. International Journal of Cancer, 2016, 139 (7): 1534 - 1545.