IRAQI Academic Scientific Journals

University of Kirkuk Journal For Administrative and Economic Science (2026) 16 (1): 513-521

UKJAES

University of Kirkuk Journal For Administrative and Economic Science

ISSN:2222-2995 E-ISSN:3079-3521

University of Kirkuk Journal For Administrative and Economic Science

Alawjar Mohammed Qasim Yahya & Al-Hasso Talal Abdul Razzaq. Bayesian Predictions of Standardized Rates of Lung Cancer Incidence by Age (A Case Study in Iraq from 2005 – 2019). *University of Kirkuk Journal For Administrative and Economic Science* (2026) 16 (1):513-521.

Bayesian Predictions of Standardized Rates of Lung Cancer Incidence by Age (A Case Study in Iraq from 2005 – 2019)

Mohammed Qasim Yahya Alawjar [1], Talal Abdul Razzaq Al-Hasso [2]

[1,2] *Department of Statistics and Informatics-College of Computer and Mathematics/University of Mosul, Mosul, Iraq*

mqy.alawjar@uomosul.edu.iq [1], talal744740@uomosul.edu.iq [2]

Abstract: Cancer is among the deadliest diseases affecting humans due to its characteristics that lead to rapid deterioration and death within a short period. Lung cancer is perhaps the most dangerous of these diseases, being a leading cause of death worldwide. This study calculated the age-based standardized rates of lung cancer incidence in Iraq during the period 2005–2019. WHO population standards were used to standardize rates across age groups, and hierarchical Poisson-Gamma was applied to address uncertainties in younger age groups. The results showed that the ASR for males increased from 107.9 to 111.6 per 100,000 people, while the results for females remained stable. 32.8 to 33.8. These results demonstrate the effectiveness of the Bayesian framework in improving the accuracy of epidemiological burden estimates.

Keywords: ASR, Bayesian, Lung Cancer, Hierarchical Poisson-Gamma, Forecasting.

التنبؤات البايزية لمعدلات الإصابة بسرطان الرئة المعيارية حسب العمر (دراسة حالة في العراق من ٢٠٠٥ - ٢٠١٩)

م. محمد قاسم يحيى الاوجار [١]، م.م. طلال عبد الرزاق الحسو [٢]

[١،٢] *قسم الاحصاء والمعلوماتية-كلية علوم الحاسوب والرياضيات/جامعة الموصل، الموصل، العراق*

mqy.alawjar@uomosul.edu.iq [1], talal744740@uomosul.edu.iq [2]

المستخلص: يُعد السرطان من أخطر الأمراض التي تصيب البشر نظرًا لخصائصه التي تؤدي إلى التدهور السريع والوفاة في غضون فترة قصيرة. وربما يكون سرطان الرئة أخطر هذه الأمراض، حيث يُعد سببًا رئيسيًا للوفاة في جميع أنحاء العالم. وقد حسبت هذه الدراسة المعدلات المعيارية القائمة على العمر لحدوث سرطان الرئة في العراق خلال الفترة ٢٠٠٥-٢٠١٩. واستُخدمت معايير السكان لمنظمة الصحة العالمية لتوحيد المعدلات عبر الفئات العمرية، وطُبقت معامل بواسون جاما الهرمي لمعالجة أوجه عدم اليقين في الفئات العمرية الأصغر. وأظهرت النتائج أن معدل الإصابة بالمرض للذكور ارتفع من ١٠٧,٩ إلى ١١١,٦ لكل ١٠٠٠٠٠ شخص، بينما ظلت نتائج الإناث مستقرة. ٣٢,٨ إلى ٣٣,٨. وتُظهر هذه النتائج فعالية الإطار البايزي في تحسين دقة تقديرات العبء الوبائي. بإيجاز الاهداف الاساسية ومجال البحث، يصف منهجية البحث المستخدمة ويلخص النتائج ويحدد الاستنتاجات الاساسية. كما يجب عدم كتابة المصادر المستخدمة داخل الملخص. اعادة صياغة بدلا من الاقتباس. استعمال صيغة

المبني للمعلوم وليس المبني للمجهول (لكن بدون استعمال الضمائر) وكذلك ينبغي تعريف الاختصارات بشكل كامل في حال ذكرها للمرة الاولى.

الكلمات المفتاحية: ASR، Bayesian، Lung cancer، hierarchical Poisson-Gamma، Forecasting.

Corresponding Author: E-mail: mqy.alawjar@uomosul.edu.iq

Introduction:

Studies in the field of lung cancer are among the most important studies in cancerous diseases, given that this type of disease is one of the deadliest for humans, with millions of new cases being recorded annually, especially in developing countries. Due to differences in population structures and variations in diagnostic estimates, age-standardized ratios (ASRs) are used to enable comparable results across time and regions. Since traditional methods (such as linear regression) estimate fixed points and do not account for uncertainty, especially in younger age groups , Bayesian modeling was used. This model provides a flexible framework for incorporating such information, reducing variance, and estimating reliability intervals that naturally reflect uncertainty. In this study, Bayesian modeling was used to obtain standardized rates of return (ASR) for the period 2005–2019. Reliability intervals were estimated, and the accuracy of the estimates was improved. A Poisson-Gamma hierarchical model was applied . A Bayesian linear trend model was also used to estimate ASR up to 2030.

1st: The Aime

1. To calculate the age-standardized rates (ASRs) for lung cancer incidence in Iraq.
2. To improve the accuracy of estimates and establish reliability intervals.
3. To predict the age-standardized rates for the period (2020–2030).

2nd: The theoretical aspect

1. Disease Incidence

The number of disease incidences in age groups a and year t is rarely large, especially in countries with limited resources. This leads to significant fluctuations in the actual specific rates[2].

$$\boldsymbol{\lambda_{a,t}} = \frac{y_{a,t}}{n_{a,t}} \qquad \ldots (1)$$

Therefore, one of the Bayesian modelling methods will be used, namely the Poisson-Gamma hierarchical model[1] [3] [5], where information is distributed among age groups and estimates are refined.

First stage:

We use the Poisson counting model.
We assume that the number of cases y_(a,t) follows a Poisson distribution.

$$\boldsymbol{y_{a,t}} \sim \textbf{Poisson}\ (\mu_{a,t}) \qquad \ldots (2)$$

$$\mu_{a,t} = n_{a,t} \cdot \lambda_{a,t} \qquad \ldots (3)$$

Where:

$n_{a,t}$: is the population in age group a and year t

$\lambda_{a,t}$: is the true specific mean, which is unknown

Step Two:

To estimate $\lambda_{a,t}$, we assume it follows a Gamma distribution as a primary distribution:

$$\boldsymbol{\lambda_{a,t}} \sim \textbf{Gamma}\ (\alpha, \beta) \qquad \ldots (4)$$

Where :

α: is the shape parameter

β: is the rate parameter
Using Bayes' formula [8].

$$p(\boldsymbol{\lambda}_{a,t} \mid \boldsymbol{y}_{a,t}) \propto p(\boldsymbol{y}_{a,t} \mid \boldsymbol{\lambda}_{a,t}) \cdot p(\boldsymbol{\lambda}_{a,t}) \quad \dots (5)$$

$$\boldsymbol{\lambda}_{a,t} \mid \boldsymbol{y}_{a,t} \sim \textbf{Gamma}\ (a + \boldsymbol{y}_{a,t}, \beta + n_{a,t}) \dots (6)$$

To estimate the parameters:

A. Initial values must be chosen.

$a = 0.001$
$\beta = 0.001$
(This means there is no pre-existing strong value.)

B. Using a random number generator to generate samples from

$$\boldsymbol{\lambda}_{a,t}^{(s)} \sim \textbf{Gamma}\ (\boldsymbol{a_{post}}\ \beta_{post}) \quad \dots (7)$$

where
S = 1, 2, …, s
Ex : S = 1000 sample

2. Calculate the post-hoc distribution for each age group and year:

$a_{post} = 0.001 + \boldsymbol{y}_{a,t}$
$\beta_{post} = 0.001 + n_{a,t}$

3. Calculating expected values (specific average) $\boldsymbol{\lambda}_{a,t}$

$$\boldsymbol{\lambda}_{a,t} = \frac{y_{a,t}}{\boldsymbol{n}_{a,t}} \quad \dots (1)$$

$$E(\boldsymbol{\lambda}_{a,t}) = \frac{a_{post}}{\beta_{post}} \quad \dots (8)$$

The Gamma distribution was used because it is compatible with the Poasson distribution, which facilitates calculations and allows for consideration of the uncertainty in qualitative rates [8].
Third stage:
For example, if we want to calculate the specific average for an age group, let's assume for the age group 50-54 years, it is 15 for females.
$y_{a,t} = 245$ Number of cases
$\boldsymbol{n}_{a,t} = 513.824$ population
$a_{post} = 0.001 + \mathbf{245} = 245.001$
$\beta_{post} = 0.001 + 513.824 = 513.825$
Next, we take a random sample of

$$\boldsymbol{\lambda}_{a,t}^{(s)} \sim \textbf{Gamma}\ (\boldsymbol{a_{post}}\ \beta_{post}) \quad \dots (9)$$

Using Python, we generate samples of:

$$\boldsymbol{\lambda}_{a,t} \sim \textbf{Gamma}\ (245.001, 513.825)$$

Then we repeat the process for each age group and year, and for all groups and years, using Python [4].
To calculate ASR, after estimating $\boldsymbol{\lambda}_{a,t}$, we calculate ASR using standard population weights (WHO) according to the following [2] [10].

$$ASR_t = \sum_{a=1}^{15} w_a \cdot \boldsymbol{\lambda}_{a,t} \quad \dots (10)$$

For each sample s, we calculate the ASR for year t.

$$\boldsymbol{ASR_t^{(s)}} = \sum_{a=1}^{15} \boldsymbol{w_a} \cdot \boldsymbol{\lambda}_{a,t}^{(s)} \quad \dots (11)$$

Where:
The period is 15 years, from 2005 to 2019
w_a: is the weight of age group a according to the WHO standard [10].
$\boldsymbol{\lambda}_{a,t}$: These are the estimates of the specific rate as shown in Table 1.

Table (1): Global population weights by age group

No.	age group	WHO
1	0–4	12
2	5–9	10
3	10–14	9
4	15–19	9
5	20–24	8
6	25–29	8
7	30–34	6
8	35–39	6
9	40–44	6
10	45–49	6
11	50–54	5
12	55–59	4
13	60–64	4
14	65–69	3
15	70+	15

The number of weights for age groups is 15 in all countries of the world.

4. Calculating the 95% Validity Intervals

After obtaining the ASRs for the period from 2005 to 2019, we find the validity intervals. We then arrange the averages in ascending order and calculate the percentages at 2.5% and 97.5%
The lower limit is= the value at 2.5%,
and the upper limit is = the value at 97.5%.

5. Predicting ASRs:

To predict the ASRs for the period from 2020 to 2030, we will use a simple linear model based on the results obtained from applying the hierarchical model [1] [10].
The state space for this model is...

$$\boldsymbol{ASR_t} \sim \mathrm{N}\,(\beta_t , \delta^2_{obs}) \quad \dots (12)$$
$$\boldsymbol{ASR_t} = \beta_0 + \beta_1 . t + \epsilon_t \quad \dots (13)$$
$$\boldsymbol{\beta_t} = \boldsymbol{\beta_{t-1}} + \boldsymbol{\delta_{t-1}} + \boldsymbol{\epsilon_t} \quad \dots (14)$$
$$\boldsymbol{\delta_t} = \boldsymbol{\delta_{t-1}} + \boldsymbol{\eta_t} \quad \dots (15)$$

Where:
β_0 : Intercept
β_1 : Slope (Direction)
$\boldsymbol{\epsilon_t} \sim \mathrm{N}\,(0, \delta^2)$: Random Error
When
$\boldsymbol{\epsilon_t} \sim \mathrm{N}\,(0, \delta^2_{level})$, $\boldsymbol{\eta_t} \sim \mathrm{N}\,(0, \delta^2_{trend})$

Prediction steps:

1. Find the trend from historical data (2005-2019) [6][9].

$$\boldsymbol{\beta_t} = \frac{\sum(t-\bar{t}).(ASR_t - \overline{ASR})}{\sum(t-\bar{t})^2} \quad \dots (16)$$

2 . For the purpose of predicting future years [7] , equation (13) is used:

$$\widehat{ASR}_t = \beta_{2019} + \beta_1 .(t - 20219) + \epsilon_t \quad \dots (17)$$
$$\widehat{ASR}_{2030} = \beta_{2019} + \beta_1 .(2030 - 20219) + \epsilon_t$$

3. To estimate uncertainty:
- Find the standard deviation of the ASR from historical data.
- Add a margin for error that increases with time.

$$= ASR_{2019} . \delta . \sqrt{1 + (t - 2019)} \quad \dots (18)$$

$\sqrt{t}\,.\,\delta\,.1.96 =$ الخطأ في ٢٠٣٠

3rd: The Applied Aspect

The following section deals with the applied aspect of this research. Data on lung cancer cases were obtained from the official health reports of the Iraqi Ministry of Health.

Bayesian modeling was applied to this data to calculate standardized rates of lung cancer cases by age group in Iraq during the period (2005-2019). WHO population standards were used to standardize the rates for different age groups. The following are the application steps:

1. Using WHO population standards to standardize the rates for different age groups
2. Data were collected for each year **t** from 2005 to 2019 for each age group a (0-4, 5-9, ..., 70-74).

$y_{a,t}$: was the number of cases, and

$n_{a,t}$: was the population.

3. The specific gravity $\boldsymbol{\lambda_{a,t}}$ was calculated using equation (1).
4. The ASR was calculated using the WHO weights according to Table (1). Using equation (3), we obtained the ASR for each year separately (2005-2019) and for each sex separately, as shown in Table (2).

Table No. (2): ASR for Males

No.	**age group**	**ASR Male**
1	0–4	29010.6
2	5–9	27003.5
3	10–14	25873.57
4	15–19	26667.22
5	20–24	32397.2
6	25–29	32623.7
7	30–34	38606.93
8	35–39	35082.41
9	40–44	42784.2
10	45–49	41780.44
11	50–54	37269.6
12	55–59	38386.8
13	60–64	39312.43
14	65–69	46093.39
15	70- 74	46055.72

Epidemiological pyramid of lung cancer in males for the year 2019

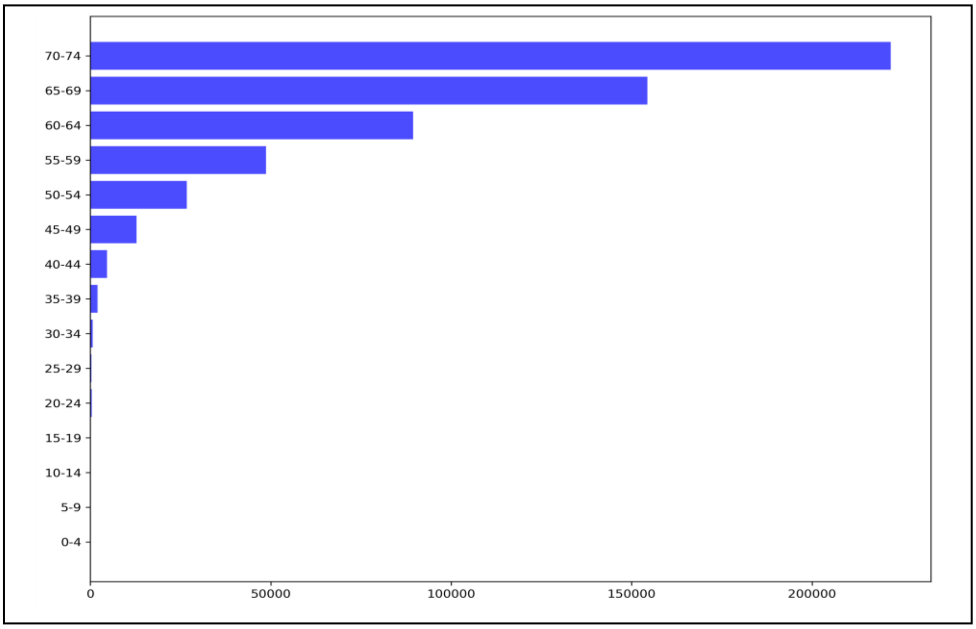

Figure (1): Specific rate per 100,000

The same applies to females; we calculate ASR as shown in Table (3) below:

Table No. (3): ASR for Females

No.	**Age group**	**Females_ASR**
1	0-4	0
2	5-9	0
3	10-14	0
4	15-19	50.16
5	20-24	329.07
6	25-29	497.71
7	30-34	143.57
8	35-39	1184.26
9	40-44	2128.98
10	45-49	5870.86
11	50-54	9987.61
12	55-59	17291.76
13	60-64	31623.36
14	65-69	44529.93
15	70-74	77959.44

Epidemiological pyramid of lung cancer in females for the year 2019

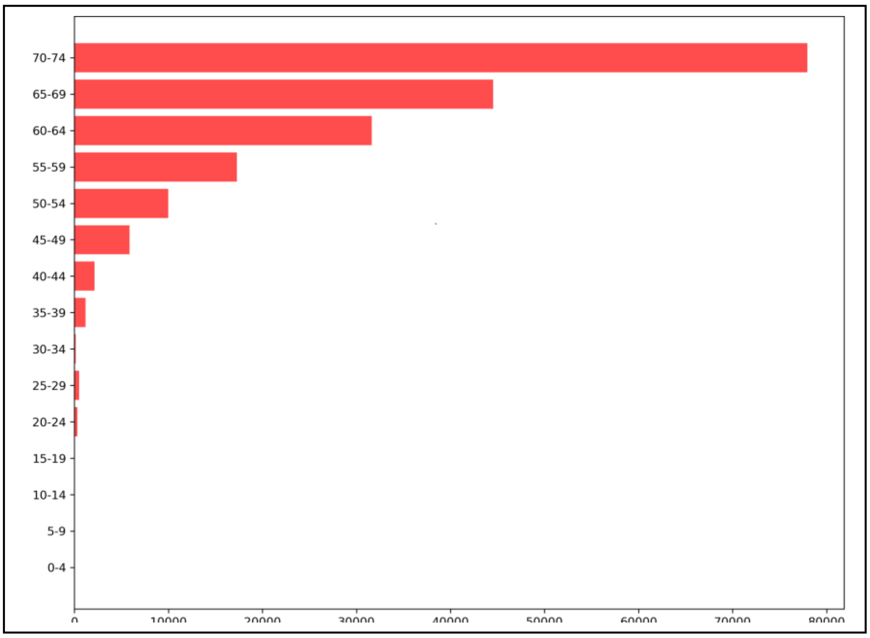

Figure (2): ASR per 100,000

5. Predicting Standard Rates for the Period (2020-2030)
Using Bayesian Modeling, we perform the following steps:

A. Assume that the number of cases is distributed according to a Poisson distribution.

B. Assume an initial distribution of type Gamma, where, from Equation No. (4), we choose initial values for the coefficients.

$a = 0.001$

$\beta = 0.001$

C. We calculate the dimensionless distribution using Bayesian equations. The dimensionless distribution is as follows: as in equation number (6).

$$\lambda_{a,t} \mid y_{a,t} \sim \text{Gamma} \left(a + y_{a,t}, \beta + n_{a,t} \right) \dots (6)$$

D. We calculate the ASR of the samples that will be taken from the dimensional distribution for each λ_(a,t) and we calculate the ASR for each sample using equation No. (11) where we can obtain samples with the number of years and the number of age groups. We obtained 1000 samples in this study, as we generated this number of random samples using Python program code.

$$\boldsymbol{ASR_t^{(s)}} = \sum_{a=1}^{15} \boldsymbol{w_a} \cdot \boldsymbol{\lambda_{a,t}^{(s)}} \qquad \ldots (11)$$

E. We estimate the reliability intervals. After calculating the $ASR_t^{(s)}$ we arrange them in ascending order. Then we calculate the limits, which are...
The lower limit is= the value at 2.5%,
and the upper limit is = the value at 97.5%.

F. Forecasting the years from 2020 to 2030
A simple linear model will be used, assuming that the ASR follows a linear trend. The parameters for equation (13) will be determined.

$$\boldsymbol{ASR_t} = \beta_0 + \beta_1 \cdot t + \epsilon_t \qquad \ldots (13)$$

Using equations number (16)

$$\boldsymbol{\beta_t} = \frac{\sum(t-\bar{t}).(ASR_t - \overline{ASR})}{\sum(t-\bar{t})^2} \qquad \ldots (16)$$

Therefore, the prediction for any year is made using equation number (17).

$$\widehat{\boldsymbol{ASR}}_t = \beta_{2019} + \beta_1 \cdot (t - 20219) + \epsilon_t \qquad \ldots (17)$$

G. Estimating uncertainty in forecasting
Using Equation No. (18)

$$\text{Error for the year 2019 t} = ASR_{2019} \cdot \delta \cdot \sqrt{(t - 2019)} \qquad \ldots (18)$$

If we want to calculate the error for the year 2030, the equation would be as follows:

$$\text{Error for the year 2030} = ASR_{2030} \cdot \delta \cdot \sqrt{(2030 - 2019)}$$

$$\text{Error for the year 2030} = ASR_{2030} \cdot \delta \cdot \sqrt{(11)}$$

Thus, we obtain the periods of reliability and uncertainty as follows.

Table No. (4): Expected ASR values for males up to 2030 with validity periods

No.	Year	ASR_Females	*Credibility_Distance _٪٩٥Lower*	*Credibility_Distance _٪٩٥Upper*
1	2020	47273.23	34350.19	60196.28
2	2021	48490.74	30214.8	66766.69
3	2022	49708.25	27324.88	72091.62
4	2023	50925.76	25079.67	76771.85
5	2024	52143.27	23246.46	81040.08
6	2025	53360.78	21705.91	85015.65
7	2026	54578.29	20387.12	88769.45
8	2027	55795.8	19243.91	92347.69
9	2028	57013.31	18244.17	95782.44
10	2029	58230.82	17364.56	99097.08
11	2030	59448.33	16587.43	102309.22

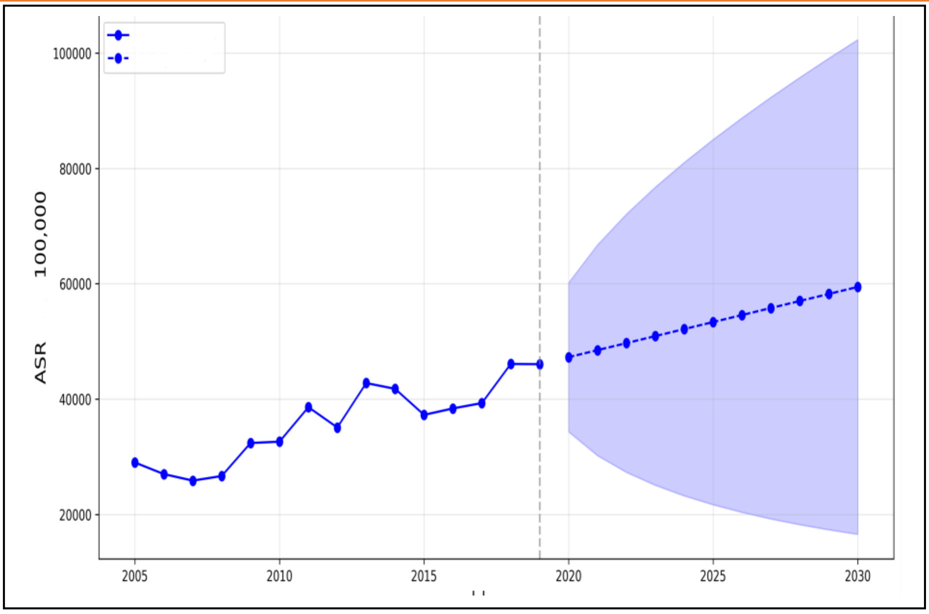

Figure (3): ASR per 100,000 Predicting ASR values for males up to 2030

Similarly, we find the expected ASR values for females for the period up to 2030.

Table No. (5): Expected ASR values for females up to 2030 with validity periods

No.	Year	ASR_Females	*Credibility_Distance _٪٩٥Lower*	*Credibility_Distance _٪٩٥Upper*
1	2020	16793.79	11132.01	22455.57
2	2021	17471.47	9464.5	25478.43
3	2022	18149.14	8342.65	27955.63
4	2023	18826.82	7503.25	30150.38
5	2024	19504.49	6844.36	32164.62
6	2025	20182.16	6313.69	34050.64
7	2026	20859.84	5880.17	35839.51
8	2027	21537.51	5523.58	37551.45
9	2028	22215.19	5229.84	39200.53
10	2029	22892.86	4988.74	40796.99
11	2030	23570.54	4792.53	42348.54

redicting ASR values for males up to 2030

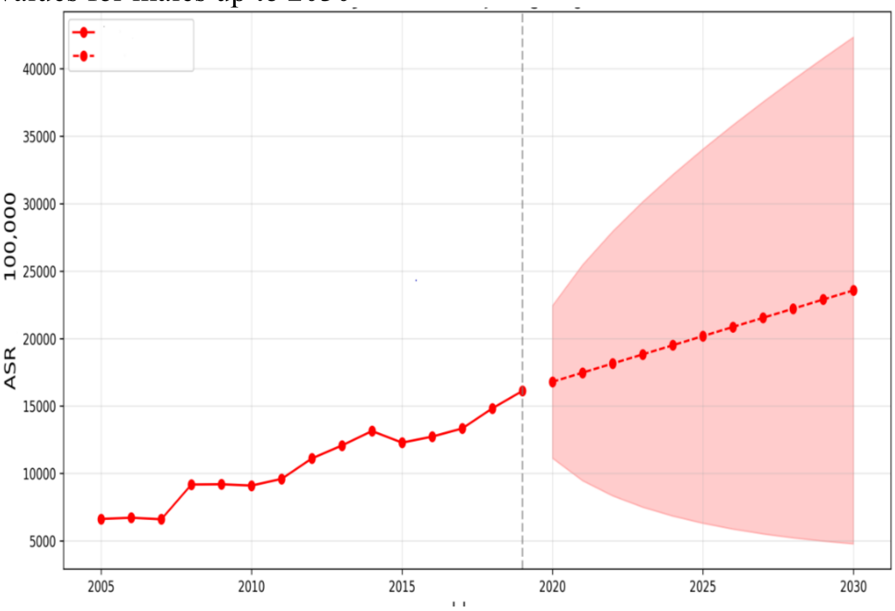

Figure (4): ASR per 100,000

Results:

The study revealed a number of findings regarding lung cancer incidence in Iraq. Among the most important of these findings are:

1- The estimated annual ASR incidence rates during the period (2005-2019) were:
- For males, the average was 107.86 per 100,000 population, showing a slight increase of 0.01 units annually.
- For females, the average was 32.76 per 100,000 population, showing relative stability with a slight increase of 0.04 units annually.

There was a significant gender gap of 3.3:1 males to females. This may be attributed to the harmful effects of smoking and air pollutants, to which males are more exposed than females.

2- Bayesian modeling reduced variance in small sample sizes and estimated 95% confidence intervals for each year. The actual results fell within these intervals with a 95% probability.

3- Using Bayesian modeling to estimate ASR up to 2030, the predictions showed the following:
In males, a projected increase of 4.1% was observed over the next decade. This may be attributed to increased risk factors such as smoking of all kinds and environmental pollution due to wars and other factors.
In females, a slight increase of 6.1% was observed, which may be due to delays in diagnosis.

4- The study results, broken down by age group, showed the following:
For males, the highest incidence rate was 221.8 per 100,000 people in the over-70 age group, and the lowest incidence rate was 0.5 per 100,000 people in the 0-4 age group. For females, the highest incidence rate is 77.9 per 100,000 people in the over-70 age group, while the lowest incidence rate is 0.0 per 100,000 people in the 0-4 age group.
This may be due to the fact that lung cancer is age-related; with cumulative exposure to risk factors, the chance of developing the disease increases.
The gender gap in older age groups is attributed to poor habits such as widespread smoking, the variety of smoking methods, and weak monitoring, follow-up, and diagnosis of health, especially among those over 50.

5- From all of this, we can conclude that males are more susceptible to this disease, largely due to the prevalence of smoking in its various forms.

Reference

1- Agresti, Alan, and David B. Hitchcock. "Bayesian inference for categorical data analysis." Statistical Methods and Applications 14, no. 3 (2005).DOI: 10.1007/s10260-005-0121-y.
2- Ala'din A, MD,. "Health in Iraq, the Current situation, our vision for the future and areas of work." Ministry of Health, 2nd edition, Baghdad, December; 2004. Available from: http://www.who.int/hac/crises/irq/background/Iraq_Health_in_Iraq_second_edition.pdf
3- Benhamou, E., Saltiel, D., Laraki, R., and Atif, J.: a conjugate prior bayesian optimization , W. (1975). The physical basis for earthquake prediction. Technology Review .
4- Cihan, P. Bayesian hyperparameter optimization of machine learning models for predicting biomass gasification gases. Applied Sciences 15. (2025). https://doi.org/10.3390/app15031018 .
5- Dada, B., Nwulu, N., and Olukanmi, S. Bayesian optimization with optuna for enhanced soil soil nutrient prediction: A comparative study with genetic algorithm and particle swarm ptimization. Smart . Agricultural Technology . (2025). https://doi.org/10.1016/j.atech.2025.101136.
6- Das, Sourish, and Dipak K. Dey. "On bayesian analysis of generalized linear models using the jacobian technique." The American Statistician 60, no. 3 (2006).
7- Dey, Dipak, Sujit K. Ghosh, and Bani K. Mallick, eds. Generalized linear models: A Bayesian perspective. Vol. 5. CRC Press, 2000.
8- Santos, M.. Bayesian optimization for hyperparameter tuning. Journal of Bioinformatics and Artificial Intelligence 2. (2022).
9- Wells, Martin T. "Generalized linear models: A Bayesian perspective." Journal of the American Statistical Association 96, no. 453 (2001) .
10-WHO. " Health topics, Cancer.". Accessed 30 September 2013 at: http://www.who.int/topics/cancer/en/