

Bayesian network identification of lung cancer-supportive behavior

Novendri Isra Asriny

2023

Background

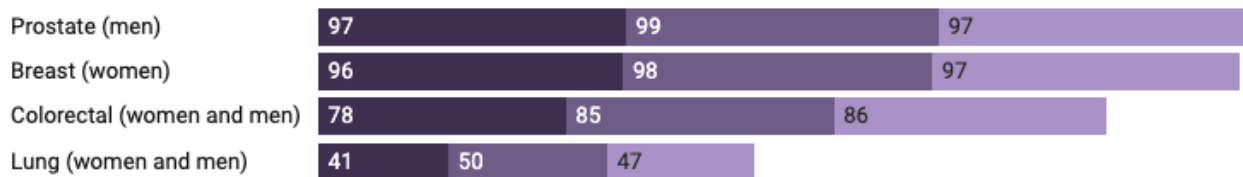
Lung cancer is the leading cause of cancer incidence and mortality worldwide, accounting for an estimated 2 million cases and 1.8 million deaths annually. Lung cancer is the second most typical cancer diagnosis for both men and women, after breast cancer and prostate cancer (Wheless et al., 2013).

According to recent UK data, the average 5-year survival rate for breast cancer, or the percentage of patients still alive after their diagnosis, is around 80 (World Cancer Research Fund International, 2020). Some malignancies, like lung cancer, have a 5-year survival rate of less than 20%.

One-year survival rates

All stages at diagnosis combined

■ UK ■ US ■ Net



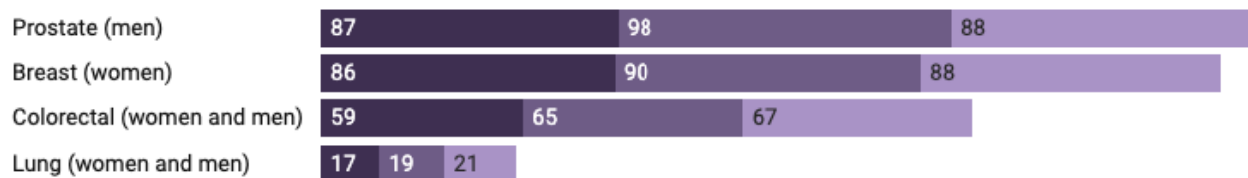
Netherlands data is for invasive breast cancer; Netherlands colorectal data is avg of colon (83) and rectum (88)

Source: World Cancer Research Fund International • Created with [Datawrapper](#)

Five-year survival rates %

All stages at diagnosis combined

■ UK ■ US ■ NL



Netherlands data is for invasive breast cancer; Netherlands colorectal data is avg of colon (66) and rectum (67)

Source: World Cancer Research Fund • Created with [Datawrapper](#)

Fig 1. Cancer Survival Data (are still alive #year after diagnosis or start of treatment)

Source : (World Cancer Research Fund International, 2020)

According to Fig. 1, lung cancer has a low survival rate in comparison to other malignancies for both men and women. This has motivated academics to investigate in greater detail which factors are causally related to the factors suggested to cause lung cancer.

The Bayesian Network technique is a mathematical model that uses conditional probabilities to show the causal connection between random variables (variables that cannot be used to establish causality). (Imoto et al., 2005).

As a beginning point for creating a recommendation system for recognizing lung cancer, it is hoped that by using the Bayesian Network approach to data on lung cancer, it will be possible to identify factors that have a high likelihood of determining the causation of the variables that are owned.

Literature Review

1. Bayesian Network

According to (Hasniati et al., 2019), the Bayesian Network technique is a probabilistic graphical model that may depict variables and the causal relationships between them. Nodes that serve as variables or initial hypotheses of a statement can be drawn as graphs in a Bayesian network. According to (Noriega et al., 2019), the Bayesian Network employs a method based on Bayes' Theorem called conditional probability, which is indicated by interpreting the likelihood of a condition A.

The formula of Bayes' Theorem is i.e (Noriega et al., 2019) :

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Informations :

$P(A|B)$ = posterior probability which is the probability that A occurs after B occurs.

$P(A \cap B)$ = the probability events B and A occur simultaneously

$P(B|A)$ = Likelihood, where the probability of B occurring after event A

$P(A)$ = priors, the probability of event A

$P(B)$ = priors, the probability of event B

2. Conditional Probability

The Parent and Child scale nodes are the two conditions in the Bayesian Network's design. The available data scales are discrete ordinal, nominal, interval, or ratio. (Laitila & Virtanen, 2022).

For any pairing of parent nodes, CPT calculates the state of the immediate predecessors and the probability distribution of the child nodes, or descendants. A proper CPT typically requires expert elicitation because the scale of measurement for ranking nodes, such as "low, medium, and high," frequently changes. Due to factors like cognitive tiredness or a lack of time, it is frequently challenging for the expert to assess all relevant variables because a single CPT can have dozens or even hundreds of components.

The formula for Conditional Probability is as follows (Rohmer, 2020):

$$\theta_{ijk} = P(X_i = k | pa(X_i) = j)$$

Informations :

X_i = parents node

$i = 1, \dots, n$

$j = 1, \dots, q_i$

$k = 1, \dots, r_i$

Methodology

1. Dataset

The dataset utilized in this study is annual updated Lung Cancer data from Kaggle with the CC BY-NC-SA 4.0 License. The source of the information used is The information is gathered from the online lung cancer prediction system website.

2. Population & Sample

The general public and global community serve as the study's population, while the survey respondents who said they had lung cancer serve as the study's sample.

3. Flow Research

In this study, the following is the flow of the research conducted.

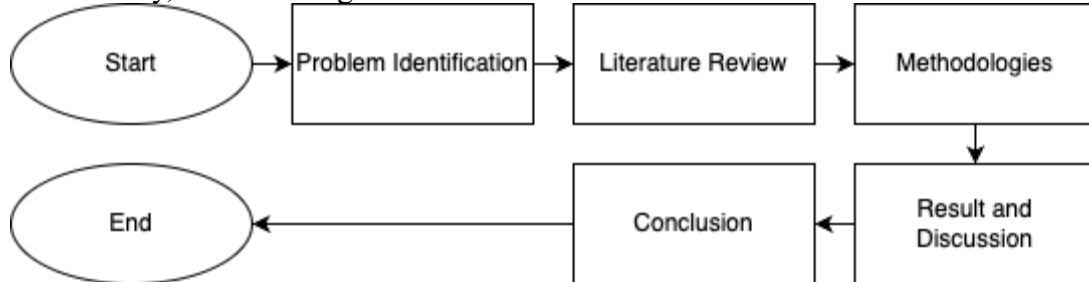


Fig 2. Flow Research

4. Variables

Table 1. Research Variables

Variables	Data Scale
Smoking	Nominal
Yellow Fingers	Nominal
Anxiety	Nominal
Peer Pressure	Nominal
Chronic Disease	Nominal
Fatigue	Nominal
Allergy	Nominal
Wheezing	Nominal
Alcohol Consuming	Nominal
Coughing	Nominal
Shortness Of Breath	Nominal
Swallowing Difficulty	Nominal
Chest Pain	Nominal
Lung Cancer Status	Nominal

5. Bayesian Network

Implementing a Bayesian Network involves the following steps:

- Construction of a Bayesian Network Structure;
- Definition of Parameters;
- Make a Conditional Probability Table

Result & Discussion

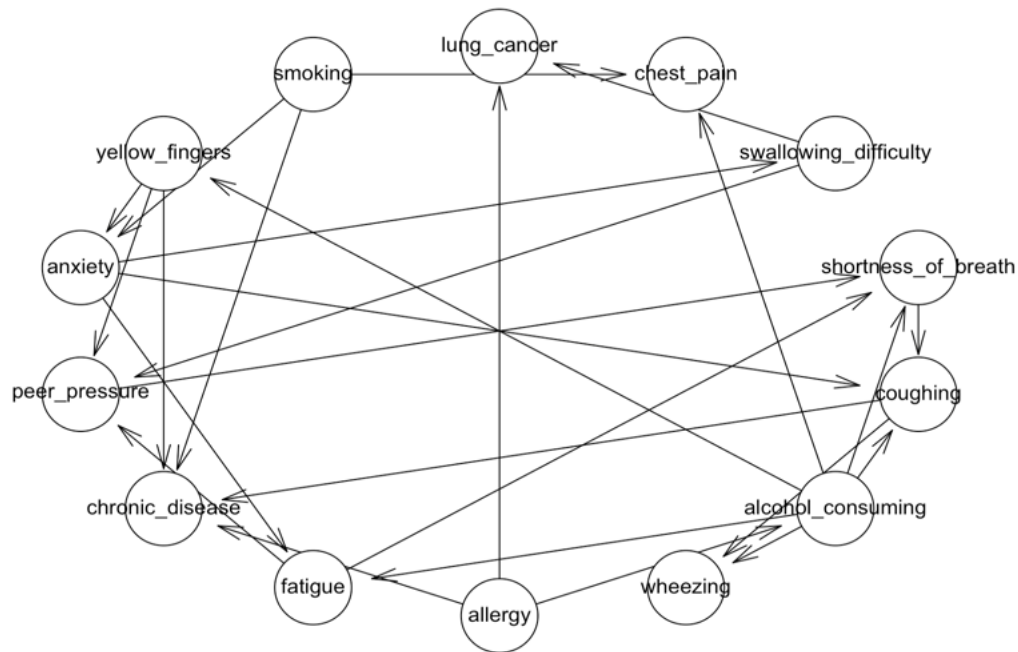


Fig 3. Networks Between Variables and/or Factor

From **Fig 3**, we can create our first hypothesis before we take a look more deep to probability, we can know if allergy and swallowing difficulty have causality to lung cancer, then smoking can make chest pain, anxiety, chronic disease, then alcohol consuming can make coughing, wheezing, chest pain, shortness of breath.

Also from **Fig 3**, we can know the journey someone indicate with lung cancer.

1. Smoking can make anxiety and then turn into swallowing difficulty and indicate to lung cancer
2. Alcohol consuming can make yellow finger, someone with yellow finger can be anxiety that makes swallowing difficulty and indicate with lung cancer
3. Allergy have direct causality to lung cancer (low indicate)

Beside that from Fig 1. We can know the journey someone indicate with chronic disease

1. Alcohol consuming can make yellow finger that will be indicate to chronic disease
2. Smoking have direct causality to chronic disease
3. Yellow finger have impact to peer pressure that will makes shortness of breath that makes coughing with the final indicate to chronic disease

Table 2. Connection Nodes From Bayesian Network

From	To
Alcohol Consuming	Chest Pain
	Coughing
	Fatigue
	Shortness of Breath
	Wheezing
	Yellow Fingers
Allergy	Alcohol Consuming
	Chronic Disease
	Lung Cancer
Anxiety	Coughing
	Fatigue
	Swallowing Difficulty
Coughing	Chronic Disease
	Wheezing
Fatigue	Peer Pressure
	Shortness of Breath
Peer Pressure	Shortness of Breath
Shortness of Breath	Coughing
Smoking	Anxiety
	Chest Pain
	Chronic Disease
Swallowing Difficulty	Lung Cancer
	Peer Pressure
Yellow Fingers	Anxiety
	Chronic Disease
	Peer Pressure

Based on **Table 2** it can be seen that most of the causes of physical or psychological illness come from Alcohol Consuming, Smoking and Yellow Fingers. Where, if a congenital disease (Allergy) places the cause also Alcohol Consuming.

Table 3. Parent and Child Node

Root	Parent	Child
Smoking	(none)	Anxiety
		Chronic Disease
		Chest Pain
Yellow Finger	Alcohol Consuming	Anxiety
		Peer Pressure
		Chronic Disease
Anxiety	Smoking	Fatigue
	Yellow Fingers	Coughing
	(none)	Swallowing Difficulty
Peer Pressure	Yellow Fingers	Shortness of Breath
	Fatigue	
	Swallowing Difficulty	
Chronic Disease	Smoking	(none)
	Yellow Fingers	
	Allergy	
	Coughing	
Fatigue	Anxiety	Peer Pressure
	Alcohol Consuming	Shortness of Breath
Allergy	(none)	Chronic Disease
		Alcohol Consuming
		Lung Cancer

Based on **Table 3** it can be seen that the parent nodes are Smoking and Allergy because these 2 root variables do not have parents. In addition, it is also known that smoking can cause anxiety, chronic disease, and chest pain. Then, Yellow Finger can cause Anxiety, Peer Pressure, and Chronic Disease. Meanwhile, Anxiety can cause Fatigue, Coughing, and Swallowing Difficulty where Anxiety can be caused by Smoking and Yellow Fingers. Then Peer Pressure can cause Shortness of Breath where Peer Pressure is caused by Yellow Fingers, Fatigue, and Swallowing Difficulty. Then in Chronic Disease it can be caused by Smoking, Yellow Fingers, Allergy, and Coughing. Furthermore, fatigue can cause Peer Pressure (this is due to a lack of desire to do the activities you once enjoyed) and Shortness of Breath where Fatigue can be caused due to Anxiety and Alcohol Consuming. Finally, Allergy can cause Chronic Disease, Alcohol Consuming, and Lung Cancer.

Table 4. Conditional Probability Tables (CPTs) of Smoking in General

Root	Conditional Probability	
	Yes	No
Smoking	44%	56%

Based on **Table 4** it can be seen that from the data used the Lung Cancer Dataset CC BY-NC-SA 4.0 the potential for patients who do not smoke is greater (56%) than those who smoke.

Table 5. Conditional Probability Tables (CPTs) of Yellow Fingers with Alcohol Consuming

Yellow Fingers	Alcohol Consuming	
	Yes	No
Yes	27%	56%
No	73%	44%

Based on **Table 5** it can be seen that those who do not have Yellow Fingers are not due to Alcohol Consuming where this is marked with a probability of 73%.

Table 6. Conditional Probability Tables (CPTs) of Yellow Fingers with Alcohol Consuming and Anxiety

Yellow Fingers (Yes)		
Anxiety	Smoking	
	Yes	No
Yes	45%	10%
No	55%	90%

Based on **Table 6** it can be seen that Yellow Fingers disease is not affected by Anxiety or Smoking which is characterized by a probability value of 90%

Table 7. Conditional Probability Tables (CPTs) of Fatigue, Swallowing Difficulty, Peer Pressure and Yellow Fingers

Fatigue (Yes), Swallowing Difficulty (Yes)		
Peer Pressure	Yellow Fingers	
	Yes	No
Yes	31%	13%
No	69%	87%
Fatigue (No), Swallowing Difficulty (No)		
Peer Pressure	Yellow Fingers	
	Yes	No
Yes	56%	74%
No	44%	26%

Based on **Table 7** it can be seen that when patients have Fatigue and Swallowing Difficulty they do not experience Peer Pressure and Yellow Fingers with a probability value of 87%. Meanwhile, when the patient does not suffer from Fatigue and Swallowing Difficulty, they experience Peer Pressure and do not experience Yellow Fingers, which has a probability value of 74%.

Table 8. Conditional Probability Tables (CPTs) of Yellow Fingers, Allergy, Coughing, Chronic Disease, and Smoking

Yellow Fingers (No), Allergy (No), Coughing (No)		
Chronic Disease	Smoking	
	Yes	No
Yes	25%	63%
No	75%	36%

Yellow Fingers (Yes), Allergy (No), Coughing (No)		
Chronic Disease	Smoking	
	Yes	No
Yes	46%	25%
No	54%	75%

Yellow Fingers (No), Allergy (Yes), Coughing (No)		
Chronic Disease	Smoking	
	Yes	No
Yes	29%	29%
No	71%	71%

Yellow Fingers (Yes), Allergy (Yes), Coughing (No)		
Chronic Disease	Smoking	
	Yes	No
Yes	75%	58%
No	25%	42%

Yellow Fingers (No), Allergy (No), Coughing (Yes)		
Chronic Disease	Smoking	
	Yes	No
Yes	64%	100%
No	36%	0%

Yellow Fingers (Yes), Allergy (No), Coughing (Yes)		
Chronic Disease	Smoking	
	Yes	No
Yes	95%	63%

No	36%	0%
Yellow Fingers (No), Allergy (Yes), Coughing (Yes)		
Chronic Disease	Smoking	
	Yes	No
Yes	0%	80%
No	100%	20%
Yellow Fingers (Yes), Allergy (Yes), Coughing (Yes)		
Chronic Disease	Smoking	
	Yes	No
Yes	32%	31%
No	68%	69%

Based on **Table 8** it can be seen that:

- When the patient does not have Yellow Fingers, Allergy and Coughing, then the patient does not have Chronic Disease but is a Smoker with a probability of 75%.
- When the patient has Yellow Fingers, does not have Allergy and Coughing, the patient does not have Chronic Disease and the patient does not smoke with a probability of 75%.
- When a patient does not have Yellow Fingers and Coughing but has an Allergy, there is a possibility of 2, the patient does not have Chronic Disease and/or a Smoking-Non Smoking Patient with a probability of 71%.
- When a patient has Yellow Fingers and Allergy but not Coughing, then the patient has Chronic Disease and is a Smoking patient with a probability value of 75%.
- When the patient does not have Yellow Fingers, Allergy but Coughing, the patient has Chronic Disease and does not smoke has a probability value of 100%.
- When a patient has Yellow Fingers and Coughing, but does not have Allergies, the patient has Chronic Disease and Smokers have a probability value of 95%
- When the patient does not have Yellow Fingers, but suffers from Allergy and Coughing, then he has Chronic Disease and is a non-smoker with a probability value of 80%. While not experiencing Chronic Disease but smokers have a 100% probability value.
- While patients who have all of them, the possibility of patients not experiencing Chronic Disease and smokers is 68% while those who do not smoke are 69%.

Table 9. Conditional Probability Tables (CPTs) of Alcohol Consuming, Fatigue and Anxiety

Alcohol Consuming (No)		
Fatigue	Anxiety	
	Yes	No
Yes	25%	21%
No	75%	79%

Alcohol Consuming (Yes)		
Fatigue	Anxiety	
	Yes	No
Yes	23%	64%
No	77%	35%

Based on **Table 9** it can be seen that patients who are not Alcohol Consuming and not Fatigue are likely to not experience Anxiety with a possible value of 79%, this is 4% greater than those experiencing Anxiety. Whereas in patients who are Alcohol Consuming without suffering from Fatigue, the possibility of patients experiencing Anxiety is 77%.

Table 10. Conditional Probability Tables (CPTs) of Coughing, Wheezing, Alcohol Consuming

Coughing (No)		
Wheezing	Alcohol Consuming	
	Yes	No
Yes	83%	44%
No	15%	56%

Coughing (Yes)		
Wheezing	Alcohol Consuming	
	Yes	No
Yes	31%	26%
No	69%	73%

Based on **Table 10** it can be seen that patients who do not have Coughing and experience Wheezing and Alcohol Consuming have a probability value of 83%. Whereas patients who have Coughing without experiencing Wheezing and Alcohol Consuming are 73%, this is 4% greater than those who do not experience Wheezing but are Alcohol Consuming.

Table 11. Conditional Probability Tables (CPTs) of Alcohol Consuming, Shortness of Breath, Coughing and Anxiety

Alcohol Consuming (Yes), Shortness of Breath (No)		
Coughing	Anxiety	
	Yes	No
Yes	34%	84%
No	66%	16%

Alcohol Consuming (Yes), Shortness of Breath (Yes)		
Coughing	Anxiety	
	Yes	No

Yes	5%	27%
No	95%	72%

Based on **Table 11** it can be seen that patients who are Alcohol Consuming and do not suffer from Shortness of Breath experience Coughing and not Anxiety have a probability value of 84%. Whereas patients who are Alcohol Consuming and suffer from Shortness of Breath do not experience Coughing but experience Anxiety with a probability value of 95%.

Table 12. Conditional Probability Tables (CPTs) of Anxiety and Swallowing Difficulty

Swallowing Difficulty	Anxiety	
	Yes	No
Yes	77%	29%
No	22%	71%

Based on **Table 12** it can be seen that patients who experience Swallowing Difficulty and Anxiety have a probability value of 77%.

Table 13. Conditional Probability Tables (CPTs) of Alcohol Consuming, Chest Pain, and Smoking

Alcohol Consuming (No)		
Chest Pain	Smoking	
	Yes	No
Yes	51%	70%
No	48%	30%

Alcohol Consuming (Yes)		
Chest Pain	Smoking	
	Yes	No
Yes	51%	12%
No	49%	88%

Based on **Table 13** it can be seen that patients who are not Alcohol Consuming but experience Chest Pain and are not smokers have a probability value of 70%.

Table 14. Conditional Probability Tables (CPTs) of Swallowing Difficulty, Chest Pain and Smoking

Swallowing Difficulty (No)		
Chest Pain	Smoking	
	Yes	No
Yes	43%	5%
No	57%	94%

Swallowing Difficulty (Yes)		
Chest Pain	Smoking	
	Yes	No
Yes	7%	0%
No	93%	100%

Based on **Table 14** it can be seen that patients who do not suffer from Swallowing Difficulty and do not experience Chest Pain or are non-smokers have a probability value of 94%. Whereas patients who suffer from Swallowing Difficulty without experiencing Chest Pain and patients who do not smoke have a probability value of 100% (7% greater than those who smoke).

Table 15. Conditional Probability Tables (CPTs) of Fatigue, Alcohol Consuming, Shortness of Breath and Peer Pressure.

Fatigue (No), Alcohol Consuming (No)		
Shortness of Breath	Peer Pressure	
	Yes	No
Yes	41%	86%
No	59%	14%

Fatigue (Yes), Alcohol Consuming (No)		
Shortness of Breath	Peer Pressure	
	Yes	No
Yes	18%	15%
No	82%	85%

Fatigue (No), Alcohol Consuming (Yes)		
Shortness of Breath	Peer Pressure	
	Yes	No
Yes	59%	80%
No	41%	19%

Fatigue (Yes), Alcohol Consuming (Yes)		
Shortness of Breath	Peer Pressure	
	Yes	No
Yes	3%	58%
No	97%	42%

Based on **Table 15** it can be seen that patients who do not suffer from Fatigue and do not experience Alcohol Consuming experience Shortness of Breath and do not experience Peer Pressure have a probability value of 86%. Meanwhile, patients who suffer from Fatigue without Alcohol Consuming do not experience Shortness of Breath and do not experience Peer Pressure,

which has a probability value of 85% (3% greater than those experiencing Peer Pressure). Meanwhile, patients who suffer from Fatigue and Alcohol Consuming do not experience Shortness of Breath but experience Peer Pressure, which has a probability value of 97%.

Table 16. Conditional Probability Tables (CPTs) of Swallowing Difficulty and Anxiety

Swallowing Difficulty	Anxiety	
	Yes	No
Yes	77%	29%
No	22%	71%

Based on **Table 16** it can be seen that patients who experience Anxiety and Swallowing Difficulty have a possible value of 77%.

Conclusion

The Bayesian Networks approach led to the following finding (high probability of causality in the research variables):

- a. Patients who do not smoke are greater (56%) than those who smoke
- b. Yellow Fingers Disease is not affected by Anxiety or Smoking which is characterized by a probability value of 90%
- c. Patients with Fatigue and Swallowing Difficulty do not experience Peer Pressure and Yellow Fingers with a probability value of 87%
- d. The patient does not have Yellow Fingers, Allergy but Coughing, so the patient has Chronic Disease and does not smoke has a probability value of 100%.
- e. Patients who are Alcohol Consuming and suffer from Shortness of Breath do not experience Coughing but experience Anxiety with a probability value of 95%
- f. Patients who suffer from Swallowing Difficulty without experiencing Chest Pain and patients who do not smoke have a 100% probability, while those who smoke are 93%
- g. Patients who do not suffer from Fatigue and do not experience Alcohol Consuming experience Shortness of Breath and do not experience Peer Pressure have a probability value of 86%
- h. Patients who suffer from Fatigue and Alcohol Consuming do not experience Shortness of Breath but experience Peer Pressure has a probability value of 97%
- i. Anxiety and Swallowing Difficulty has a probability value of 77%.

References

- Hasniati, H., Arianti, A., & Philip, W. (2019). Penerapan Metode Bayesian Network Model Pada Sistem Diagnosa Penyakit Sesak Nafas Bayi. *IKRA-ITH INFORMATIKA : Jurnal Komputer Dan Informatika*, 3(2), 19–26. <https://journals.upi-yai.ac.id/index.php/ikraith-informatika/article/view/314>
- Imoto, S., Matsuno, H., & Miyano, S. (2005). Gene Networks: Estimation, Modeling, and Simulation. *Computational Systems Biology*, 2000, 205–228. <https://doi.org/10.1016/B978-012088786-6/50030-7>
- Laitila, P., & Virtanen, K. (2022). Advancing construction of conditional probability tables of Bayesian networks with ranked nodes method. *International Journal of General Systems*, 51(8), 758–790. <https://doi.org/10.1080/03081079.2022.2086541>
- Noriega, A. R., Ramírez, R. J., Tapia, J. J., Castillo, V. H., & Jimenez, S. (2019). Construction of conditional probability tables of Bayesian networks using ontologies and Wikipedia. *Computacion y Sistemas*, 23(4), 1275–1289. <https://doi.org/10.13053/CyS-23-4-2705>
- Rohmer, J. (2020). Uncertainties in conditional probability tables of discrete Bayesian Belief Networks: A comprehensive review. *Engineering Applications of Artificial Intelligence*, 88. <https://doi.org/10.1016/j.engappai.2019.103384>
- Wheless, L., Brashears, J., & Alberg, A. J. (2013). Epidemiology of lung cancer. *Lung Cancer Imaging*, 1–15. https://doi.org/10.1007/978-1-60761-620-7_1
- World Cancer Research Fund International. (2020). *Cancer Survival Statistics*. <https://www.wcrf.org/Cancer-Trends/Cancer-Survival-Statistics/>