

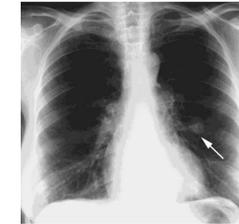
# A Comparison of Prominent Models to Predict the Malignancy of Solitary Pulmonary Nodules



By: Lilyana Gross

Advisor: Dr. Judith Canner  
California State University, Monterey Bay

## X-Ray Examples of SPNs <sup>6</sup>



## Introduction

I investigated and compared commonly used stochastic models to predict if SPNs (Solitary Pulmonary Nodules) found in patients are malignant or benign.

### Lung Cancer

- One of the leading causes of cancer deaths since 1985<sup>4</sup>
- Manifests in the appearance of solitary pulmonary nodules (SPNs) in the lungs, which can either be diagnosed as **malignant** (cancerous) or **benign** (non-cancerous)
- Early detection of malignant SPNs can increase survival rate of cancer patients by 40%<sup>4</sup>

Mathematical models serve as tools for physicians to refrain from performing unnecessary procedures in patients whose SPNs are most likely benign.

## The Models

**The Mayo Model 1<sup>1</sup>:** A model created in 1999 by the Mayo Clinic used clinical data to predict the outcome of SPNs

**The Mayo Model 2<sup>2</sup>:** The researchers at the Mayo Clinic studied the outcomes of SPNs predicted by physicians. A curve was created but no predictive model.

**The Li-Wang Model<sup>3</sup>:** In 2012, the Li-Wang model included family history of cancer as a predictor variable.

**The Wang Model<sup>4</sup>:** In 2013, Wang continued his research and added biomarker data and used an advanced statistical technique to develop the model.

All models used the same basic logistic regression model:

$$p = \frac{e^x}{1 + e^x}$$

Where  $p$  is the probability a SPN is malignant, and  $e$  is the base of the natural logarithm.

The factors (independent covariates) that influence each model are listed in Table 1 with their respective  $\beta$  values. The  $\beta$  values can be interpreted as the weight each variable holds as a predictor of cancerous SPNs.

## Model Comparison

### Covariates Tested in each Model (Table 1):

Model	$\beta_0$	Age	Smoker	Past Cancer	Diameter	Spiculation	Upper Lobe	Family History	Calcification	Clear Border	CEA	CYFRA 21-1	Satellite Lesion
Mayo 1	-6.827	0.039	0.792	1.339	0.127	1.041	0.7838						
Li-Wang	-4.496	0.070			0.676	0.736		1.267	-1.615	-1.408			
Wang	-4.294	0.035	1.029		0.633	2.027	2.673	0.974	-3.295	-1.631	0.221	0.200	-1.923

**Sensitivity Analysis:** Used to quantify the robustness and uncertainty of model

Model	Sensitivity	Specificity
Mayo 1	Not Provided	Not Provided
Mayo 2	Not Provided	Not Provided
Li-Wang	94.5%	70%
Wang	90.7%	81.2%

\*Sensitivity: Probability the model will produce a **malignant prediction when the SPN truly is cancerous**

\*Specificity: Probability the model will produce a **true benign prediction when the SPN truly is not cancerous**

Model	Area Under the Curve (AUC)
Mayo 1	0.798 ± 0.047
Mayo 2	0.835 ± 0.033
Li-Wang	0.874 ± 0.028
Wang	0.935 ± 0.011

\*Area Under the Characteristic Curve quantifies how accurate the model diagnosis SPNs. An AUC closer to 1 implies a stronger model.

**Cost Exploration :** Used to assess the accessibility and monetary restrictions of the model

### How accessible is the model?

The Mayo Model 1 is accessible online. →

The Mayo Model 2 is the predictions of physicians without mathematical computations.

The Li-Wang Model and Wang Model are public, but must be computed manually.

An open source tool for Mayo Model 1 online at [www.medscape.com](http://www.medscape.com).

### What is the cost to obtain the data?

- The clinical information necessary for The Mayo Models 1 & 2, and the Li-Wang can be collected by easily by physicians.
- The Wang Model requires biomarker series tests and CT scans, which raises the cost of obtaining the data

### Do physicians have the computational abilities?

The Wang Model requires a LASSO technique that requires a highly trained statistician or computer scientist.

## Conclusion

**Recommendation:** Based on the sensitivity analysis and cost exploration, I recommend the Wang Model. Though the Mayo Model 1 is the most accessible, The Wang Model has the highest specificity and AUC. It is crucial to obtain true negatives when early detection can increase life expectancy. If the data collection is not possible for the Wang Model, I recommend the Li-Wang Model over The Mayo Model 1.

**Future Work:** Though all of these models aid in the diagnosis of SPNs, they are merely a tool to aid physicians. Further research needs to be done to make The Wang Model more accessible to all physicians and patients. More research needs to be done to find additional predictor variables that maximize the predictive power of stochastic models that predict the malignancy of SPNs.

## References

[1] Swensen SJ1, Silverstein MD, Edell ES, Trastek VF, Aughenbaugh GL, Ilstrup DM, Schleck CD. (1991) Solitary pulmonary nodules: clinical prediction model versus physicians. *Mayo Clin Proc.* 1999 Apr;74(4):319-29. [2] Li, Y., Wang, J. (2012). A mathematical model for predicting malignancy of solitary pulmonary nodules. *World Journal of Surgery*, 36(4), 830-835. [3] Dong, J., Sun, N., Li, J., Liu, Z.,

Zhang, B., et al. (2014). Development and validation of clinical diagnostic models for the probability of malignancy in solitary pulmonary nodules. *Thoracic Cancer*, 5(2), 162-168. [4] Wang, J., Wu, H., Sun, T., Li, X., Wang, W., et al. (2013). Prediction models for solitary pulmonary nodules based on curvelet textural features and clinical parameters. *Asian Pacific Journal of Cancer Prevention*: APJCP, 14(10), 6019. [5] s, E., Demanet, L., Donoho, D., Ying, L. (2006). Fast discrete curvelet transforms. *Multiscale Modeling and Simulation*, 5(3), 861. [6] a, N., Kahn, (2010) Lung hamartoma imaging. *Radiology*: 13(2), 111-117

## Acknowledgements

I would like to thank Dr. Judith Canner for taking me on as her mentee for my capstone research. Additional acknowledgements to the faculty, staff, and students in the Mathematics and Statistics Department. Without all of you this work would not have been possible.