

---

# **RESEARCH REPORT**

---

Serial No. 464

March 2009

PREDICTING LYMPH NODE INVOLVEMENT: A SIMPLE  
STATISTICAL FORMULA

by

Howell Tong



**THE UNIVERSITY OF HONG KONG**  
**DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE**

# PREDICTING LYMPH NODE INVOLVEMENT: A SIMPLE STATISTICAL FORMULA

Howell Tong

London School of Economics and Political Science & University of Hong Kong

## Summary

This note provides a simple formula to calculate the probability of nodal involvement in invasive breast cancer in women after obtaining a clear sample of axillary lymph nodes. Two tables are provided to help medical practitioners. The formula is generic and can be applied to other medical conditions. The note also gives a simple formula which approximates the exact probability closely in many practical situations and has the very simple interpretation as the product of false negative probability and prior odds.

*Key words: axillary sampling, breast cancer, conditional probability, false negative probability, nodal involvement, prior odds.*

## 1. Introduction

It is well known that, internationally speaking, breast cancer is the most common cancer among women and is the greatest cause of cancer deaths. In 2000 alone, more than one million women were diagnosed (22% of all female cancer diagnoses) and 373 000 women died (14% of all cancer deaths among women) of breast cancer.<sup>1</sup> The pathologic status of the axilla plays a significant role in accurate prognosis for breast cancer patients. Suppose that from the axillary tissue provided by the surgeon only 3 lymph nodes are found and they all turn out to be clear. Should a further sample be taken?

In Forrest<sup>2</sup>, 3 or 4 nodes were used and a false negative probability of 10% was claimed. Later Forrest *et al.*<sup>3</sup> reported the results of a study conducted in Cardiff involving 417 patients concerning the impact of 4-node sample on breast cancer management. Since then the method has apparently become one of the standard procedures in the UK, Denmark and perhaps elsewhere. Now, how closely should we adhere to the four-node-axillary-sampling scheme? Answers to this question as well as the earlier question and other similar questions are relevant in breast cancer management. For this purpose, we need the probability of axillary lymph node involvement after an n-node sample has been declared negative.

---

<sup>1</sup> Parkin DM, Bray FI, Devesa SS. (2001). Cancer burden in the year 2000. The global picture. *Eur J Cancer*, **37**, S4–S66.

<sup>2</sup> Forrest, A.P.M., Stewart, H.J., Roberts, M.M. and Steele, R.J.C. (1982). Simple mastectomy and axillary node sampling (pectoral node biopsy) in the management of primary breast cancer. *Ann. Surg.*, **196**, 371–377.

<sup>3</sup> Forrest, A.P.M., Everington, D, McDonald, C., Steele, R.J.C., Chetty, U. and Stewart, H.J. (1995). The Edinburgh randomized trial of axillary sampling or clearance after mastectomy. *Brit. J Surg*, **82**, 1504–1508.

## 2. A simple formula

For events  $A$  and  $B$ , let  $Pr(A)$  and  $Pr(A/B)$  denote respectively the probability of  $A$  and the conditional probability of  $A$  given  $B$ . Let  $NI$  denote the event of nodal involvement and  $n-$  denote the event that a sample of  $n$  nodes is declared negative. Thus,  $Pr(n-|NI)$  is the false negative probability, i.e. the probability that a sample of  $n$  axillary lymph nodes is declared negative while there is in fact nodal involvement. Before axillary sampling, the oncologist/physician usually has some idea, one way or the other, of the prior probability,  $Pr(NI)$ , of nodal involvement. This information can be utilized and its value can be updated after axillary sampling to what is called the posterior probability of nodal involvement and is denoted by  $Pr(NI|n-)$ . Specifically, the updating formula is as follows (assuming that  $Pr(n-) > 0$  and  $Pr(NI) > 0$ ).

$$\Pr(NI|n-) = \Pr(n-|NI) \times \Pr(NI) / K,$$

where  $K = Pr(n-) = [1 - Pr(NI)] \times 1 + Pr(NI) \times Pr(n-|NI)$ . To prove the updating equation, we only need to notice that by the definition of conditional probability

$$\Pr(NI|n-) \times Pr(n-) = \Pr(n-|NI) \times \Pr(NI) = \Pr(NI \text{ and } n-),$$

and

$$\Pr(n-) = \Pr(n-|NI) \times \Pr(NI) + \Pr(n-|\text{not } NI) \times \Pr(\text{not } NI),$$

where  $Pr(n-|\text{not } NI)$  is clearly equal to 1 and  $Pr(\text{not } NI) = 1 - Pr(NI)$ .

Denote  $Pr(NI) / Pr(\text{not } NI)$  as the prior odds of nodal involvement. If  $Pr(n-|NI) \times$  prior odds is smaller than one, which is often the case, the posterior probability may be closely approximated by the following formula.

$$\Pr(NI|n-) \sim \Pr(n-|NI) \times \text{Prior odds.}$$

**In words, the posterior probability of nodal involvement is approximately equal to the product of the false negative probability and the prior odds (if the product is less than one).**

From this approximate value, we can get the exact value very quickly: we only need to divide the approximate value by 1 plus the approximate value.

## 3. Implementation of the formula

To implement the formula, we need some reliable results for both the false negative probability,  $Pr(n-|NI)$ , and the prior probability of nodal involvement,  $Pr(NI)$ . For the

false negative probability, Tanaka *et al.*<sup>4</sup> conducted a study in 2006 on 237 primary breast cancer patients at stages I and II prospectively and found  $Pr(n-/NI)$  to be about 7% when  $n = 4$ . A separate prospective study was reported by Cserni<sup>5</sup> based on records and slides of patients who all had axillary clearance. He ranked the nodes of each patient in descending order by size and firmness. He then observed that by looking at the first 3 nodes, the first 4 nodes, the first 5 nodes and the first 6 nodes, the false negative probabilities are 11%, 6%, 5% and 3% respectively. Note the similarity between Cserni's result and the result obtained by Tanaka *et al.* above for the case  $n = 4$ . In what follows, we make the assumption that the surgeon usually removes the most palpable nodes using his/her experience. For the prior probability, the medical practitioner can, of course, use his/her own experience to assign a value between 0 and 1. As one of many possibilities, we give Table 1 below, which is based on the three-variable model fitted by Olivotto *et al.*<sup>6</sup>, with a minor modification in order to avoid a prior probability of value 0 or 1.

**Table 1: Prior probability, Pr(NI), of axillary nodal involvement before axillary sampling, expressed in terms of number of breast cancer patients per 100\*.**  
(LVI=lymphovascular invasion)

	Tumour Size (in mm.)							
	0~5	6~10	11~15	16~20	21~25	26~30	31~50	51~100
LVI absent								
Can you feel the tumour?								
No	4	11	12	10	20	33	57	67
Yes	6	13	19	20	30	27	34	65
Yes; axillary nodes too	50	75	93	80	70	71	77	93
LVI present								
Can you feel the tumour?								
No	5	18	30	39	42	33	50	95
Yes	23	28	37	49	65	67	66	77
Yes; axillary nodes too	98	98	98	92	90	96	98	98

<sup>4</sup> Tanaka, K., Yamamoto, D., Kanematsu, S., Okugawa, H. and Kamiyama, Y. (2006). A four node axillary sampling trial on breast cancer patients. *Breast*. **15**, 203-209.

<sup>5</sup> Cserni, G. (1999). The reliability of sampling three to six nodes for staging breast cancer. *J. Clinical Pathology*, **52**, 681-683.

<sup>6</sup> Olivotto, I. A., Jackson, J.S.H., Mates, D., Andersen, S., Davidson, W., Bryce, C.J. and Ragaz, J. (1998). Prediction of axillary lymph node involvement of women with invasive breast carcinoma. *Cancer*, **83**, 948-955.

\* The cell entries can also be read as percentages so that typically the first cell means that the prior probability is 4%.

For example, a female patient without LVI who can feel her tumour, which turns out to be of size 18mm, is given a prior probability of 20%, by Table 1, of having nodal involvement. This figure is an initial estimate and is by its nature quite rough. Note that the prior odds is thus  $20\% \div 80\%$ , i.e.  $\frac{1}{4}$ . We shall come back to this prior odds later.

To refine the estimate, we can apply the updating formula derived above. Table 2 below is constructed to aid the medical practitioner. For example, for the female patient in the above example, if only 3 nodes have been removed and they turn out to be negative, then her posterior probability of nodal involvement is given by Table 2 to be 3%. We can, without using Table 2, arrive at practically the same figure by applying the approximate formula above: we note that the false negative probability is in this case 11% and the prior odds is  $\frac{1}{4}$ , giving a product of 2.75%, or 3% after rounding.

Now, the posterior probability drops to 1% if, in fact she has a negative 5-node sample, which could arise, for example, as a result of the pathologist, say, subsequently finding two more lymph nodes in the tissue and they turn out to be negative too.

As a further example, Table 2 shows that the posterior probabilities of nodal involvement for a patient with prior probability of 10% are all equal to 1%, if her sample of  $n$  nodes is declared negative, for  $n = 3, 4$  and  $5$ . This information might be useful to the medical team, for example, to decide how many lymph nodes to be removed, bearing in mind the possible physical and psychological side effects on the patient caused by their removal. The general idea is that a smaller surgical cut is less likely to cause side effects than a bigger cut.

**Table 2: Posterior probability,  $\Pr(\text{NI}|n-)$ , of axillary nodal involvement after axillary sampling, expressed in terms of number of breast cancer patients per 100.**

	<i>Number of nodes sampled and found to be clear</i>			
	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>
<i>Prior probability of axillary nodal involvement expressed as number of patients per 100</i>				
<b>5</b>	1	0	0	0
<b>10</b>	1	1	1	0
<b>15</b>	2	1	1	1
<b>20</b>	3	2	1	1
<b>25</b>	4	2	2	1
<b>30</b>	5	3	2	1
<b>35</b>	6	4	3	2
<b>40</b>	7	4	3	2
<b>45</b>	8	5	4	2
<b>50</b>	10	7	5	3
<b>55</b>	12	8	6	4
<b>60</b>	14	10	7	4
<b>65</b>	17	12	8	5
<b>70</b>	20	14	10	7
<b>75</b>	25	17	13	8
<b>80</b>	31	22	17	11
<b>85</b>	38	28	22	15
<b>90</b>	50	39	31	21
<b>95</b>	68	57	49	36
<b>100</b>	100	100	100	100

## 4. Conclusions

It is hoped that the simple formula developed in this note will help medical practitioners in their breast cancer management. It should help them in addressing issues such as the optimal size of the axillary sample, as well as highlighting the useful role of a well informed prior probability. Of course, if an ideal medical test is widely available which has zero false negative probability, then the formula will not be necessary because zero false negative probability implies zero posterior probability.

The formula can also be applied to a general medical condition, as long as the false negative probability is known about the test designed to test the presence of the condition. Denoting the medical condition by  $M$  and a negative result of the test by  $T^-$ , we only need to replace  $NI$  by  $M$  and  $n^-$  by  $T^-$  in the formula. The formula can be easily extended to cover the case of two statistically independent tests by treating the posterior probability from one test as the prior probability for the other. The order of the test is immaterial as far as probability calculation is concerned. Medical considerations might dictate the order, however. Clearly the same argument can be extended to cover the case of more than two statistically independent tests.

**Acknowledgements:** I thank Dr. Michaël Selby and Mrs. Mary Tong for encouragement and suggestions and the University of Hong Kong for partial support under their Distinguished Visiting Professorship scheme.

(19<sup>th</sup> April 2009)

## Current Departmental Research Reports

<u>Serial No.</u>	<u>Date</u>	<u>Research Report Title</u>	<u>Author(s)</u>
452	Jun-07	Asymptotics for ruin probabilities of a two-dimensional renewal risk model with heavy-tailed claims	Yiqing Chen, Kam C. Yuen and Kai W. Ng
453	Jul-07	Semiparametric model based inference in the presence of missing responses	Qihua Wang and Pengjie Dai
454	Sep-07	A threshold approach for peaks-over-threshold modelling using maximum product of spacings	T.S.T. Wong and W.K. Li
455	Nov-07	On a dynamic mixture GARCH model	Xixin Cheng, Philip L.H. Yu and W.K. Li
456	Mar-08	Statistical inference in partially-varying-coefficient single-index model	Qihua Wang and Liugen Xue
457	Mar-08	Robustness diagnosis for bootstrap inference	W.L. Lok and Stephen M.S. Lee
458	Oct-08	Decay parameter and invariant measures for Markovian bulk-arrival queues with control at idle time	Junping Li and Anyue Chen
459	Oct-08	Decay properties and quasi-stationary distributions for stopped Markovian bulk-arrival and bulk-service queues	Anyue Chen, Junping Li and Zhang Hanjun
460	Dec-08	Uniqueness and extinction properties of interacting branching collision processes	Anyue Chen, Junping Li and Zhang Hanjun
461	Dec-08	A geometric process $\delta$ -shock maintenance model	Yeh Lam
462	Jan-09	A resampling-based test for threshold structure in time series	Guodong Li and Wai Keung Li
463	Mar-09	A multiple imputation approach for clustered interval-censored survival data	K.F. Lam, Ying Xu and Tak-Lun Cheung
464	Mar-09	Predicting lymph node involvement: A simple statistical formula	Howell Tong