Prevalence of cancer and optimal cut-off levels for mathematical models to distinguish between benign and malignant adnexal masses

[1]Anneleen Daemen, [2,3]Caroline Van Holsbeke, [4]Robert Fruscio, [5]Stefano Guerriero, [6]Artur Czekierdowski, [7]Lil Valentin, [8]Luca Savelli, [9]Antonia C Testa, [10]Nicoletta Colombo, [2,11]Tom H Bourne, [2]Ignace Vergote, [1]Bart De Moor, [2]Dirk Timmerman

[1]Department of Electrical Engineering ESAT/SCD, Katholieke Universiteit Leuven, Leuven, Belgium; [2]Department of Obstetrics and Gynecology, University Hospitals Leuven, campus Gasthuisberg, Katholieke Universiteit Leuven, Leuven, Belgium; [3]Ziekenhuis Oost-Limburg, Genk, Belgium; [4]San Gerardo Hospital, Monza, Italy; [5]Ospedale San Giovanni di Dio, Cagliari, Italy; [6]Medical University, Lublin, Poland; [7]University Hospital, Malmö, Sweden; [8]Reproductive Medicine Unit, Bologna, Italy; [9]Università Cattolica del Sacro Cuore, Rome, Italy; [10]Prof.ssa Ginecologic Oncology Unit, Division of Gynecology, IEO, Milano, Italy; [11]St Georges Hospital Medical School, London, UK

Running title: Prevalence of cancer and optimal cut-off levels

Correspondence:

A. Daemen

Department of Electrical Engineering

Katholieke Universiteit Leuven

Kasteelpark Arenberg 10 bus 2446

3001 Leuven

Belgium

e-mail: anneleen.daemen@esat.kuleuven.be

Key words:

Statement of translational relevance

Abstract

Purpose

Two logistic regression models LR1 and LR2 to distinguish between benign and malignant adnexal masses were developed in phase 1 of a multicenter study by the International Ovarian Tumor Analysis (IOTA) group. The goal of this retrospective analysis is to verify whether the models perform differently according to prevalence of malignancy and whether the cut-off levels of the models require alteration per center or type of center.

Experimental design

19 centers participated in this study and contributed 1938 new cases. Concerning the types, a distinction was made according to the prevalence of malignant cases into centers with a prevalence of less than 15%, between 15 and 30% and above 30%. The area under the ROC curves (AUC) were compared using bootstrapping. The optimal cut-off level per center and type was chosen corresponding to a sensitivity level as high as possible (preferable above 90%) while still keeping a good specificity (80%).

Results

Both LR1 and LR2 performed (statistically) better in centers with a lower prevalence of malignant cases. The AUC of centers with less than 15% of malignancy was 0.968 and 0.950, for LR1 and LR2 respectively, 0.947 and 0.926 for centers with prevalence between 15 and 30%, and 0.938 and 0.919 for centers with more than 30% malignancies. This decrease in performance was mainly due to the decrease in specificity from over 90% to around 76%. The

optimal cut-off per center varied between 0.05 and 0.20, but the performance in the centers with a higher percentage of malignant cases did not improve by choosing a different cut-off level.

Conclusions

The performance of the logistic regression models increases with decreasing prevalence of malignancy, due to a difficult mix of benign and borderline tumors in centers with a high prevalence of malignancy. Because new cut-off levels per center would be based on 9 to 252 patients and the cut-off of 0.10 is optimal for all three types of center, it seems reasonable to use this cut-off in all centres.

Introduction

In the first phase of the study of the International Ovarian Tumor Analysis (IOTA) group, 1066 patients were gathered from 9 centers in 5 countries. Two logistic regression models LR1 and LR2 were developed using ultrasound variables to preoperatively distinguish benign from malignant adnexal masses (Timmerman *et al.*, 2005). The output of the models is the probability on malignancy, for which the cut-off to distinguish benign from malignant was set to a probability of 0.10.

In phase 2, 1938 new cases were gathered from 19 centers in 8 countries. This data set contains 542 malignant masses (28%) and 1396 benign masses (72%). Eight centers were already involved in phase 1 whereas 12 centers were new. The centers were subdivided into three types, according to the prevalence of malignancy. In 5 centers less than 15% of the cases were malignant, 6 centers had a prevalence of malignancy between 15 and 30% whereas in 8 centers prevalence of malignancy was higher than 30% (Table 1).

A retrospective study is performed to verify whether 1) the logistic regression models LR1 and LR2 perform differently between the types of center and 2) the cut-off levels of LR1 and LR2 require alteration per center or type of center.

Materials and Methods

To verify whether the logistic regression models differ in performance according to the type of center, receiver operator characteristics (ROC) curves were compared. Common methods for the comparison of the area under the ROC curves (AUC) (e.g. DeLong and DeLong) are not applicable because in this set-up the same logistic regression model needs to be compared on different sets of samples instead of different models on the same set of samples. Therefore a bootstrap-based method was used. Bootstrapping is a procedure that

involves choosing random elements with replacement from a data set. Because each element is returned to the data set after sampling, a particular element from the original data set can appear multiple times in a given bootstrap sample. The number of elements in each bootstrap sample equals the number of elements in the original data set.

The bootstrap method of Johnson (Johnson, 2001) was used. First, the original data set was split into data of 1 center type and data of the remaining centers. Each of the logistic regression models were applied separately on both subsets and a difference score x was defined as the difference between the two corresponding AUC values. Secondly, both subsets were sampled B times with replacement, with B set to 2000. The AUC for both subsets were calculated for each bootstrap sample, resulting in B difference scores $xB=(x_1, \ldots, x_B)$. Thirdly, a bootstrap confidence interval was calculated as $(2x - xB(97.5),\ 2x - xB(2.5))$ with $xB(2.5)$ an estimate of the 2.5th percentile of the population difference score and $xB(97.5)$ of the 97.5th percentile. To ascertain statistically significant differences in performance of the models between the considered center type and all other centers, the percentile confidence interval of the difference score was compared with zero. When this interval does not include zero, there is a significantly difference in performance whereas the performance differs not statistically significant when the interval includes zero.

The cut-off levels were redefined per center and center type, requiring a sensitivity as high as possible (>90%) with a good specificity (around 80%) because it is considered very important to correctly identify a malignant case (Timmerman *et al.*, 2005)

Results

The performance of the logistic regression models LR1 and LR2 are shown per type of center in Tables 2 and 3. Figures 1 and 2 show the corresponding ROC curves. For both models the AUC decreases with increasing prevalence of malignancy. To verify whether the

AUC value of each center type differs significantly from the other centers a bootstrap-based method was applied. Tables 2 and 3 show the difference score in AUC with its corresponding 95% confidence interval. Logistic regression model LR1 has a significantly better performance in the centers with a prevalence of malignancy below 15%. This observation was not statistically significant when using LR2. The decrease in performance when prevalence of malignancy exceeds 30% is not statistically significant for both models.

In Tables 4 to 9, the sensitivity, the specificity, and the accuracy are tabulated against the use of different probability levels per type of center and for both LR1 and LR2. At a cut-off of 0.10, the sensitivity ranges from 89.9 to 93.3% and fulfills the specifications. The specificity varies between 75.5 and 91.2%. For the logistic regression model LR1, the prevalence of malignancy does not greatly influence the sensitivity (90.6% - 93.3%). However, the specificity decreases from 90.6 to 77.0% when the prevalence of malignancy increases. The same holds for LR2 containing half of the ultrasound variables: the sensitivity barely varies (89.8% - 90.6%) whereas the specificity decreases from 90.1 to 75.7% with increasing prevalence of malignant masses in the centers.

To clarify the decrease in specificity, we considered how the mix of pathologies differs in the three types of center, as shown in Table 10. In centers with a higher prevalence of malignancy, the pathologies endometrioma, hydrosalpinx, and salpingitis seem to occur less, while there are more teratoma, mucinous cystadenoma, borderline, primary invasive, and metastatic masses. To know whether these pathologies cause the decrease in specificity, we studied the misclassifications in each type of center. Tables 11 and 12 show per center type the number and fraction of misclassifications, for LR1 and LR2, respectively. For both models, more simple cyst, parasalpingeal cyst, and mucinous cystadenoma masses are misclassified in centers with a prevalence of malignancy above 30%. In these centers, the

percentage of misclassified borderline masses is almost double, while less primary invasive and metastatic masses are classified wrongly.

When considering the sensitivity, the specificity, and the accuracy for different probability levels per center (results not shown), the optimal cut-off according to the above specifications – when able to obtain – fluctuated between 0.05 and 0.20. However, these results are based on 9 to 252 patients whereas the cut-off of 0.10 was based on the training set of IOTA phase 1 (i.e. 746 patients, 70% of 1066 patients). Therefore, because the cut-off of 0.10 seems to be optimal for each type of center and to keep the usability of the models as general as possible, no new cut-off values are determined per center.

Discussion

The performance of the logistic regression models decreases with increasing prevalence of malignancy. This reduction in performance is almost completely caused by a decreasing specificity at a cut-off probability of 0.10. In centers with a higher prevalence of malignancy, the mix of patients is different because more difficult patients are referred to such centers (i.e. the proportion of difficult tumors increases with experience; there are more referred patients from regional hospitals). These centers not only see more primary invasive and metastatic masses, but also more teratoma, mucinous cystadenoma, and borderline tumors. The cut-offs levels for the logistic regression models have been chosen to obtain a good sensitivity such that malignant masses are certainly classified as such. Centers with difficult benign and borderline masses therefore have a lower specificity due to misclassification of these masses as malignant.

References

1. Timmerman D, Testa A, Bourne T, Ferrazzi E, Ameye L, Konstantinovic M, Van Calster B, Collins W, Vergote I, Van Huffel S, Valentin L. Logistic regression model to distinguish between the benign and malignant adnexal mass before surgery: a multicenter study by the international ovarian tumor analysis group. Journal of Clinical Oncology 2005;23(34):8794-8801.

2. DeLong and DeLong

3. Johnson. An introduction to the bootstrap. Teaching Statistics 2001;23(2):49-54.

Table 1: Contribution and characteristics of the 19 centers involved

Center	Code	N	N malignant	N benign	% malignant	% benign	Center type
Bologna, Italy	BIT	135	11	124	8.15	91.85	1
Ontario, Canada	OCA	12	1	11	8.33	91.67	1
Milan, Italy	MIT	50	5	45	10.00	90.00	1
Cagliari, Italy	SIT	154	20	134	12.99	87.01	1
Genk, Belgium	GBE	200	27	173	13.50	86.50	1
Total center type 1		*551*	*64*	*487*	*<15%*	*>85%*	
Lund, Sweden	LSW	38	7	31	18.42	81.58	2
Milan, Italy	VIT	21	4	17	19.05	80.95	2
Malmo, Sweden	MSW	137	27	110	19.71	80.29	2
Naples, Italy	NIT	64	13	51	20.31	79.69	2
Monza, Italy	OIT	251	52	199	20.72	79.28	2
Beijing, China	BCH	73	16	57	21.92	78.08	2
Total center type 2		*584*	*119*	*465*	*15-30%*	*70-85%*	
Lublin, Poland	LPO	154	53	101	34.42	65.58	3
London, UK	KUK	65	25	40	38.46	61.54	3
Leuven, Belgium	LBE	252	97	155	38.49	61.51	3
Udine, Italy	UDI	17	7	10	41.18	58.82	3
Naples, Italy	GIT	9	5	4	55.56	44.44	3
Rome, Italy	RIT	122	68	54	55.74	44.26	3
Milan, Italy	CIT	94	53	41	56.38	43.62	3
Prague, Czech Rep.	PCR	90	51	39	56.67	43.33	3
Total center type 3		*803*	*359*	*444*	*>30%*	*<70%*	

Table 2: Bootstrap results for LR1

Center type	N	AUC (SE) center type	AUC (SE) other centers	Difference score (95% confidence interval)
1	551	0.968 (0.0078)	0.943 (0.0064)	-0.0255 ([-0.0458; -0.0065])
2	584	0.947 (0.0113)	0.952 (0.0058)	0.0054 ([-0.0226; 0.0293])
3	803	0.938 (0.0083)	0.957 (0.0073)	0.0192 ([-0.0025; 0.0410])

AUC = area under the ROC curve; SE = standard error

Table 3: Bootstrap results for LR2

Center type	N	AUC (SE) center type	AUC (SE) other centers	Difference score (95% confidence interval)
1	551	0.950 (0.0147)	0.925 (0.0078)	-0.0254 ([-0.0595; 0.0032])
2	584	0.926 (0.0142)	0.936 (0.0076)	0.0104 ([-0.0221; 0.0402])
3	803	0.919 (0.0100)	0.938 (0.0103)	0.0187 ([-0.0094; 0.0465])

AUC = area under the ROC curve; SE = standard error

Table 4: Classification of malignant and benign tumors by different probability levels using LR1 in centers with prevalence of malignancy < 15% (center type = 1)

Probability level (P)	Correctly classified		Incorrectly classified		Sensitivity (%)	Specificity (%)	Accuracy (%)	PPV (%)	NPV (%)
	Malignant	Benign	Malignant as benign	Benign as malignant					
0.01	64	216	0	271	100.0	44.4	50.8	19.1	100.0
0.05	62	399	2	88	96.9	81.9	83.7	41.3	99.5
0.10	58	444	6	43	90.6	91.2	91.1	57.4	98.7
0.15	56	457	8	30	87.5	93.8	93.1	65.1	98.3
0.20	55	462	9	25	85.9	94.9	93.8	68.8	98.1
0.25	53	469	11	18	82.8	96.3	94.7	74.6	97.7
0.30	53	470	11	17	82.8	96.5	94.9	75.7	97.7
0.35	49	470	15	17	76.6	96.5	94.2	74.2	96.9
0.40	48	471	16	16	75.0	96.7	94.2	75.0	96.7
0.45	44	471	20	16	68.8	96.7	93.5	73.3	95.9
0.50	42	471	22	16	65.6	96.7	93.1	72.4	95.5
0.55	39	474	25	13	60.9	97.3	93.1	75.0	95.0
0.60	38	477	26	10	59.4	97.9	93.5	79.2	94.8
0.65	35	477	29	10	54.7	97.9	92.9	77.8	94.3
0.70	31	477	33	10	48.4	97.9	92.2	75.6	93.5
0.75	28	481	36	6	43.8	98.8	92.4	82.4	93.0
0.80	26	482	38	5	40.6	99.0	92.2	83.9	92.7
0.85	22	485	42	2	34.4	99.6	92.0	91.7	92.0
0.90	19	486	45	1	29.7	99.8	91.7	95.0	91.5

PPV, positive predictive value; NPV, negative predictive value

Table 5: Classification of malignant and benign tumors by different probability levels using LR1 in centers with prevalence of malignancy between 15 and 30% (center type = 2)

Probability level (P)	Correctly classified		Incorrectly classified		Sensitivity (%)	Specificity (%)	Accuracy (%)	PPV (%)	NPV (%)
	Malignant	Benign	Malignant as benign	Benign as malignant					
0.01	118	120	1	345	99.2	25.8	40.8	25.5	99.2
0.05	114	337	5	128	95.8	72.5	77.2	47.1	98.5
0.10	111	383	8	82	93.3	82.4	84.6	57.5	98.0
0.15	109	399	10	66	91.6	85.8	87.0	62.3	97.6
0.20	104	411	15	54	87.4	88.4	88.2	65.8	96.5
0.25	99	418	20	47	83.2	89.9	88.5	67.8	95.4
0.30	98	432	21	33	82.4	92.9	90.8	74.8	95.4
0.35	94	438	25	27	79.0	94.2	91.1	77.7	94.6
0.40	90	442	29	23	75.6	95.1	91.1	79.6	93.8
0.45	86	445	33	20	72.3	95.7	90.9	81.1	93.1
0.50	81	447	38	18	68.1	96.1	90.4	81.8	92.2
0.55	81	448	38	17	68.1	96.3	90.6	82.7	92.2
0.60	72	452	47	13	60.5	97.2	89.7	84.7	90.6
0.65	69	455	50	10	58.0	97.8	89.7	87.3	90.1
0.70	65	458	54	7	54.6	98.5	89.6	90.3	89.5
0.75	57	459	62	6	47.9	98.7	88.4	90.5	88.1
0.80	49	460	70	5	41.2	98.9	87.2	90.7	86.8
0.85	41	462	78	3	34.5	99.4	86.1	93.2	85.6
0.90	30	463	89	2	25.2	99.6	84.4	93.8	83.9

PPV, positive predictive value; NPV, negative predictive value

Table 6: Classification of malignant and benign tumors by different probability levels using LR1 in centers with prevalence of malignancy > 30% (center type = 3)

Probability level (P)	Correctly classified		Incorrectly classified		Sensitivity (%)	Specificity (%)	Accuracy (%)	PPV (%)	NPV (%)
	Malignant	Benign	Malignant as benign	Benign as malignant					
0.01	357	128	2	316	99.4	28.8	60.4	53.0	98.5
0.05	347	293	12	151	96.7	66.0	79.7	69.7	96.1
0.10	332	342	27	102	92.5	77.0	83.9	76.5	92.7
0.15	322	362	37	82	89.7	81.5	85.2	79.7	90.7
0.20	313	382	46	62	87.2	86.0	86.6	83.5	89.3
0.25	307	392	52	52	85.5	88.3	87.0	85.5	88.3
0.30	301	397	58	47	83.8	89.4	86.9	86.5	87.3
0.35	294	403	65	41	81.9	90.8	86.8	87.8	86.1
0.40	281	412	78	32	78.3	92.8	86.3	89.8	84.1
0.45	267	417	92	27	74.4	93.9	85.2	90.8	81.9
0.50	254	422	105	22	70.8	95.0	84.2	92.0	80.1
0.55	243	427	116	17	67.7	96.2	83.4	93.5	78.6
0.60	233	431	126	13	64.9	97.1	82.7	94.7	77.4
0.65	217	435	142	9	60.4	98.0	81.2	96.0	75.4
0.70	202	436	157	8	56.3	98.2	79.5	96.2	73.5
0.75	177	438	182	6	49.3	98.6	76.6	96.7	70.6
0.80	150	440	209	4	41.8	99.1	73.5	97.4	67.8
0.85	119	440	240	4	33.1	99.1	69.6	96.7	64.7
0.90	84	443	275	1	23.4	99.8	65.6	98.8	61.7

PPV, positive predictive value; NPV, negative predictive value

Table 7: Classification of malignant and benign tumors by different probability levels using LR2 in centers with prevalence of malignancy < 15% (center type = 1)

Probability level (P)	Correctly classified		Incorrectly classified		Sensitivity (%)	Specificity (%)	Accuracy (%)	PPV (%)	NPV (%)
	Malignant	Benign	Malignant as benign	Benign as malignant					
0.01	64	72	0	415	100.0	14.8	24.7	13.4	100.0
0.05	59	383	5	104	92.2	78.6	80.2	36.2	98.7
0.10	58	439	6	48	90.6	90.1	90.2	54.7	98.7
0.15	55	453	9	34	85.9	93.0	92.2	61.8	98.1
0.20	53	462	11	25	82.8	94.9	93.5	67.9	97.7
0.25	52	467	12	20	81.3	95.9	94.2	72.2	97.5
0.30	48	468	16	19	75.0	96.1	93.6	71.6	96.7
0.35	45	471	19	16	70.3	96.7	93.6	73.8	96.1
0.40	45	472	19	15	70.3	96.9	93.8	75.0	96.1
0.45	40	473	24	14	62.5	97.1	93.1	74.1	95.2
0.50	37	475	27	12	57.8	97.5	92.9	75.5	94.6
0.55	34	477	30	10	53.1	97.9	92.7	77.3	94.1
0.60	30	482	34	5	46.9	99.0	92.9	85.7	93.4
0.65	29	484	35	3	45.3	99.4	93.1	90.6	93.3
0.70	28	484	36	3	43.8	99.4	92.9	90.3	93.1
0.75	27	485	37	2	42.2	99.6	92.9	93.1	92.9
0.80	24	486	40	1	37.5	99.8	92.6	96.0	92.4
0.85	16	486	48	1	25.0	99.8	91.1	94.1	91.0
0.90	13	486	51	1	20.3	99.8	90.6	92.9	90.5

PPV, positive predictive value; NPV, negative predictive value

Table 8: Classification of malignant and benign tumors by different probability levels using LR2 in centers with prevalence of malignancy between 15 and 30% (center type = 2)

Probability level (P)	Correctly classified		Incorrectly classified		Sensitivity (%)	Specificity (%)	Accuracy (%)	PPV (%)	NPV (%)
	Malignant	Benign	Malignant as benign	Benign as malignant					
0.01	117	68	2	397	98.3	14.6	31.7	22.8	97.1
0.05	111	325	8	140	93.3	69.9	74.7	44.2	97.6
0.10	107	382	12	83	89.9	82.2	83.7	56.3	97.0
0.15	102	402	17	63	85.7	86.5	86.3	61.8	95.9
0.20	99	409	20	56	83.2	88.0	87.0	63.9	95.3
0.25	95	420	24	45	79.8	90.3	88.2	67.9	94.6
0.30	90	428	29	37	75.6	92.0	88.7	70.9	93.7
0.35	84	433	35	32	70.6	93.1	88.5	72.4	92.5
0.40	80	436	39	29	67.2	93.8	88.4	73.4	91.8
0.45	76	443	43	22	63.9	95.3	88.9	77.6	91.2
0.50	72	447	47	18	60.5	96.1	88.9	80.0	90.5
0.55	66	452	53	13	55.5	97.2	88.7	83.5	89.5
0.60	58	454	61	11	48.7	97.6	87.7	84.1	88.2
0.65	51	457	68	8	42.9	98.3	87.0	86.4	87.0
0.70	43	458	76	7	36.1	98.5	85.8	86.0	85.8
0.75	41	461	78	4	34.5	99.1	86.0	91.1	85.5
0.80	33	462	86	3	27.7	99.4	84.8	91.7	84.3
0.85	25	464	94	1	21.0	99.8	83.7	96.2	83.2
0.90	14	464	105	1	11.8	99.8	81.8	93.3	81.5

PPV, positive predictive value; NPV, negative predictive value

Table 9: Classification of malignant and benign tumors by different probability levels using LR2 in centers with prevalence of malignancy > 30% (center type = 3)

Probability level (P)	Correctly classified		Incorrectly classified		Sensitivity (%)	Specificity (%)	Accuracy (%)	PPV (%)	NPV (%)
	Malignant	Benign	Malignant as benign	Benign as malignant					
0.01	356	59	3	385	99.2	13.3	51.7	48.0	95.2
0.05	341	284	18	160	95.0	64.0	77.8	68.1	94.0
0.10	325	336	34	108	90.5	75.7	82.3	75.1	90.8
0.15	315	357	44	87	87.7	80.4	83.7	78.4	89.0
0.20	309	374	50	70	86.1	84.2	85.1	81.5	88.2
0.25	299	385	60	59	83.3	86.7	85.2	83.5	86.5
0.30	289	391	70	53	80.5	88.1	84.7	84.5	84.8
0.35	280	396	79	48	78.0	89.2	84.2	85.4	83.4
0.40	270	405	89	39	75.2	91.2	84.1	87.4	82.0
0.45	267	413	92	31	74.4	93.0	84.7	89.6	81.8
0.50	252	419	107	25	70.2	94.4	83.6	91.0	79.7
0.55	239	425	120	19	66.6	95.7	82.7	92.6	78.0
0.60	217	429	142	15	60.4	96.6	80.4	93.5	75.1
0.65	195	435	164	9	54.3	98.0	78.5	95.6	72.6
0.70	172	438	187	6	47.9	98.6	76.0	96.6	70.1
0.75	144	438	215	6	40.1	98.6	72.5	96.0	67.1
0.80	126	442	233	2	35.1	99.5	70.7	98.4	65.5
0.85	100	442	259	2	27.9	99.5	67.5	98.0	63.1
0.90	64	443	295	1	17.8	99.8	63.1	98.5	60.0

PPV, positive predictive value; NPV, negative predictive value

Table 10: Mix of patients per type of center

		Center type 1		Center type 2		Center type 3	
	N	N	%	N	%	N	%
Benign	***1396***	***487***	***34.89***	***465***	***33.31***	***444***	***31.81***
Endometrioma	400	159	39.75	143	35.75	98	24.50
Teratoma	226	67	29.65	67	29.65	92	40.71
Simple cyst + parasalpingeal cyst	131	55	41.98	39	29.77	37	28.24
Functional cyst	77	31	40.26	21	27.27	25	32.47
Hydrosalpinx + salpingitis	49	22	44.90	15	30.61	12	24.49
Peritoneal pseudocyst	11	5	45.45	2	18.18	4	36.36
Abscess	24	5	20.83	8	33.33	11	45.83
Fibroma	81	29	35.80	22	27.16	30	37.04
Serous cystadenoma	236	66	27.97	99	41.95	71	30.08
Mucinous cystadenoma	138	40	28.99	40	28.99	58	42.03
Rare benign	18	7	38.89	5	27.78	6	33.33
Uterine Fibroid	5	1	20	4	80	-	0
Borderline	***111***	***11***	***9.91***	***28***	***25.23***	***72***	***64.86***
Borderline serous stage I	55	6	10.91	15	27.27	34	61.82
Borderline serous stage II	3	1	33.33	2	66.67	-	0
Borderline serous stage III	8	-	0	2	25	6	75
Borderline mucinous stage I	41	4	9.76	8	19.51	29	70.73
Borderline mucinous stage IV	1	-	0	1	100	-	0
Borderline endometroid stage I	1	-	0	-	0	1	100
Rare Borderline	2	-	0	-	0	2	100
Primary invasive	***373***	***45***	***12.06***	***79***	***21.18***	***249***	***66.76***
Primary invasive stage I	70	10	14.29	14	20	46	65.71
Primary invasive stage II	30	2	6.67	12	40	16	53.33
Primary invasive stage III	202	19	9.41	39	19.31	144	71.29
Primary invasive stage IV	30	7	23.33	1	3.33	22	73.33
Rare primary invasive	41	7	17.07	13	31.71	21	51.22
Metastatic	***58***	***8***	***13.79***	***12***	***20.69***	***38***	***65.52***

Table 11: Number of misclassifications per center type by LR1

	Center type 1		Center type 2		Center type 3	
	N	%	N	%	N	%
Benign	***33***	***67.35***	***66***	***73.33***	***98***	***75.97***
Endometrioma	13	26.53	21	23.33	29	22.48
Teratoma	3	6.12	11	12.22	13	10.08
Simple cyst + parasalpingeal cyst	1	2.04	3	3.33	6	4.65
Functional cyst	1	2.04	2	2.22	3	2.33
Hydrosalpinx + salpingitis	-	0	3	3.33	5	3.88
Peritoneal pseudocyst	-	0	-	0	1	0.78
Abscess	-	0	1	1.11	-	0
Fibroma	1	2.04	8	8.89	6	4.65
Serous cystadenoma	10	20.41	7	7.78	19	14.73
Mucinous cystadenoma	3	6.12	6	6.67	12	9.30
Rare benign	1	2.04	-	0	2	1.55
Uterine Fibroid	-	0	4	4.44	2	1.55
Borderline	***2***	***4.08***	***4***	***4.44***	***10***	***7.75***
Borderline serous stage I	2	4.08	3	3.33	6	4.65
Borderline serous stage II	-	0	-	0	-	0
Borderline serous stage III	-	0	-	0	-	0
Borderline mucinous stage I	-	0	1	1.11	4	3.10
Borderline mucinous stage IV	-	0	-	0	-	0
Borderline endometroid stage I	-	0	-	0	-	0
Rare Borderline	-	0	-	0	-	0
Primary invasive	***11***	***22.45***	***15***	***16.67***	***15***	***11.63***
Primary invasive stage I	3	6.12	5	5.56	3	2.33
Primary invasive stage II	-	0	1	1.11	2	1.55
Primary invasive stage III	6	12.24	7	7.78	7	5.43
Primary invasive stage IV	-	0	-	0	1	0.78
Rare primary invasive	2	4.08	2	2.22	2	1.55
Metastatic	***3***	***6.12***	***5***	***5.56***	***6***	***4.65***

Table 12: Number of misclassifications per center type by LR2

	Center type 1		Center type 2		Center type 3	
	N	%	N	%	N	%
Benign	***38***	***70.37***	***73***	***76.84***	***109***	***76.76***
Endometrioma	11	20.37	21	22.11	34	23.94
Teratoma	5	9.26	16	16.84	14	9.86
Simple cyst + parasalpingeal cyst	2	3.70	4	4.21	9	6.34
Functional cyst	2	3.70	2	2.11	3	2.11
Hydrosalpinx + salpingitis	3	5.56	3	3.16	5	3.52
Peritoneal pseudocyst	-	0	-	0	1	0.70
Abscess	-	0	-	0	1	0.70
Fibroma	2	3.70	7	7.37	7	4.93
Serous cystadenoma	10	18.52	10	10.53	20	14.08
Mucinous cystadenoma	1	1.85	6	6.32	11	7.75
Rare benign	2	3.70	-	0	2	1.41
Uterine Fibroid	-	0	4	4.21	2	1.41
Borderline	***2***	***3.70***	***3***	***3.16***	***10***	***7.04***
Borderline serous stage I	2	3.70	2	2.11	6	4.23
Borderline serous stage II	-	0	-	0	-	0
Borderline serous stage III	-	0	-	0	-	0
Borderline mucinous stage I	-	0	1	1.05	4	2.82
Borderline mucinous stage IV	-	0	-	0	-	0
Borderline endometroid stage I	-	0	-	0	-	0
Rare Borderline	-	0	-	0	-	0
Primary invasive	***11***	***20.37***	***13***	***13.68***	***17***	***11.97***
Primary invasive stage I	4	7.41	4	4.21	4	2.82
Primary invasive stage II	-	0	1	1.05	2	1.41
Primary invasive stage III	4	7.41	6	6.32	9	6.34
Primary invasive stage IV	-	0	-	0	1	0.70
Rare primary invasive	3	5.56	2	2.11	1	0.70
Metastatic	***3***	***5.56***	***6***	***6.32***	***6***	***4.23***

Figure 1: ROC curves of LR1 per center type

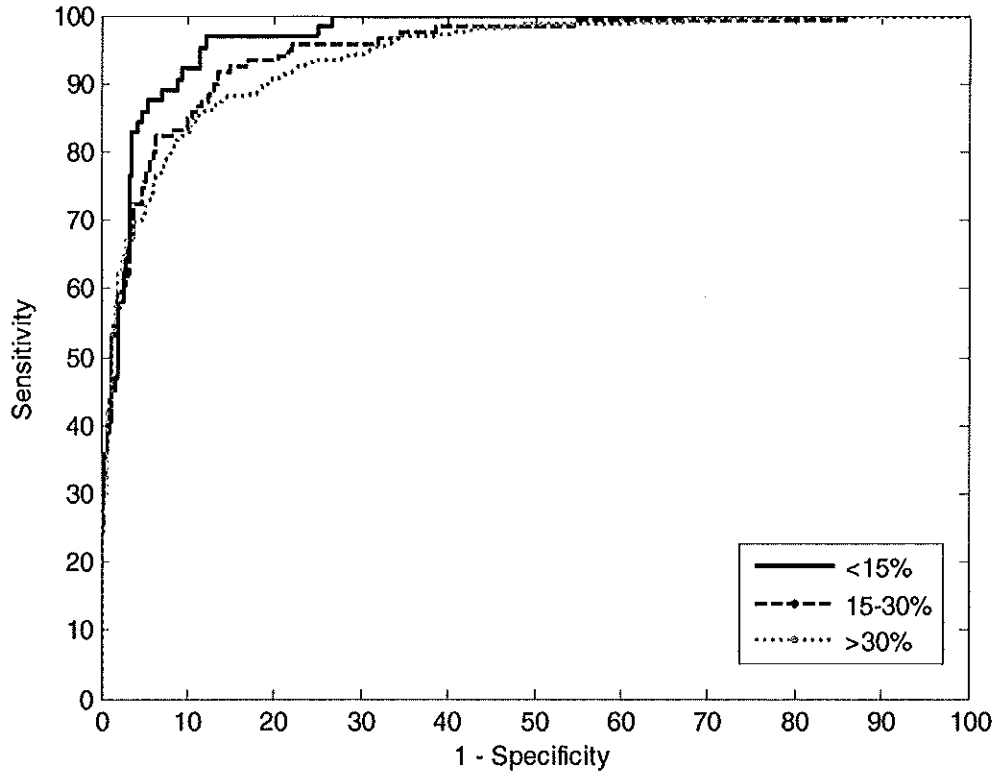

Figure 2: ROC curves of LR2 per center type