

갑상선 결절의 초음파 평가에 대한
관찰자 간의 변이성

연세대학교 대학원
의학과
최 선 형

갑상선 결절의 초음파 평가에 대한 관찰자간의 변이성

지도교수 김 은 경

이 논문을 석사 학위논문으로 제출함

2008년 06월

연세대학교 대학원

의 학 과

최 선 형

최선행 의 석사 학위논문을 인준함

심사위원 _____ 인

심사위원 _____ 인

심사위원 _____ 인

연세대학교 대학원

2008년 06월

감사의 글

논문의 주제를 정하는 것부터 마무리까지 세심한 지도를 해주신 김은경 교수님께 먼저 감사의 말씀을 전합니다.

아울러 바쁘신 와중에도 저의 학위논문을 위해 204장이나

되는 슬라이드를 여러 번 검토해 주신 곽진영, 김민정,

손은주 선생님, 통계 자문에 적극적으로 도와주신

임상시험센터의 강대룡 교수님, 연구 계획서 심사부터

공개 발표에서 아낌없는 조언을 해주신 영상의학과 김기황

교수님과 내과 이현철 교수님께도 감사 드립니다.

마지막으로 한 동안 논문 때문에 뒷순위에 밀렸던

사랑하는 우리 아들 주헌이와 가족 여러분, 고맙습니다.

저자 씀

<차례>

국문요약	8
I. 서론	10
II. 재료 및 방법	11
1. 대상군의 선택	11
2. 초음파 영상의 검토	12
3. 통계 분석	14
III. 결과	15
1. 관찰자 사이와 관찰자 내의 변이성	15
2. 관찰자 개인 판정의 정확성	19
IV. 고찰.....	20
V. 결론	25
참고문헌.....	26
영문요약.....	29

그림 차례

Fig1. A 46-year-old woman with a papillary carcinoma.....	22
Fig2. A 18-year-old woman with a papillary carcinoma.....	22
Fig3. A 55-year-old woman with multifocal papillary carcinomas.	23

표 차례

Table1. List of Ultrasonographic Finding Descriptors for Thyroid Nodule and Assessment.....	13
Table2. Inter-observer Variability in Thyroid Nodule Descriptors and Assessment Category.....	16
Table3. Intra-observer Variability in Thyroid Nodule Descriptors and Assessment Category.....	19
Table 4. .Individual Data Correlation 1 st Final Assessment with Pathology.....	20

국문 요약

갑상선 결절의 초음파 평가에 대한 관찰자간의 변이성

갑상선 결절은 흔한 내과적 질환으로 그의 일차적 진단 기기는 초음파 검사이다. 그러나, 초음파 검사는 검사 수행자에 매우 의존적인 검사로 수행자에 따른 편차가 크다. 본 연구에서는 갑상선 결절의 초음파 검사에서 영상의학과 전문의들 사이의 결절의 모양 기술 및 최종 평가의 판정 일치도와 각 전문의의 개인별 판정 일치도 및 정확도를 알아보고자 한다.

144명의 환자에서 204개의 갑상선 결절을 판정대상으로 하였으며, 전체 204개의 결절 중 악성은 89예, 양성은 115예였다. 5년 이상의 전문의 경력을 보유한 4명의 영상의학과 전문의가 4-6주 간격을 두고 독립적으로 두 번 관찰을 하였으며, 결절의 에코, 조성, 경계, 방위, 석회화, 혈류에 대한 기술 및 최종 판정을 하였다. 관찰자 간, 관찰자 내의 판정 일치도는 Cohen's kappa statistics를 이용하였으며, 최종 판정의 정확도를 계산하였다.

관찰자 간의 일치도는 에코가 약간의 일치도($\kappa=0.34$), 조성, 경계, 석회화 및 최종 판정이 보통의 일치도($\kappa=0.59$, $\kappa=0.42$, $\kappa=0.58$ 그리고 $\kappa=0.54$)를 보였으며, 방위 및 혈류가 상당한 일치도($\kappa=0.61$, $\kappa=0.64$)를 보였다. 관찰자 내의 일치도는 거의 대부분이 $\kappa>0.61$ 의 상당한

일치도를 보였다. 네 명의 관찰자의 평균 민감도와 특이도, 양성 예측도, 음성 예측도 및 정확도는 각각 88.2%, 78.7%, 76.2%, 89.6% 그리고 82.8%였다.

결론적으로 숙련된 영상학과 전문의들의 갑상선 결절의 기술 및 평가는 대부분 보통 이상의 일치도를 보였으며, 결절에 대한 최종 판정 역시 상당히 정확하였다.

핵심 되는 말: 갑상선 결절, 초음파, 변이성

갑상선 결절의 초음파 평가에 대한 관찰자간의 변이성

<지도교수 김은경>

연세대학교 대학원 의학과

최 선 형

I. 서 론

갑상선 결절은 약 19%에서 67%의 유병률을 보이는 흔한 내과적 질환이나 이들 중 만져지는 결절은 전체 인구의 4% - 7%이며, 영상 검사, 특히 초음파 검사에서 발견된 증상이 없는 우연종이 대부분을 차지한다^{1, 2}. 갑상선 결절은 여성과 노인 및 요오드 섭취 결핍 지역에게 더 흔하며, 최근에는 초음파 기기의 발달로 인해 1cm 미만의 작은 무증상 결절, 우연종의 발견이 더더욱 증가하였다. 그러나 무증상 갑상선 우연종에서 악성 위험도는 약 0.45% - 13% (mean \pm SD, 3.9% \pm 4.1%)에 불과하며 대부분의 갑상선 결절은 양성이다¹. 이러한 무증상의 갑상선 우연종 및 만져지는 갑상선 결절에 대한 진단 및 접근에 있어 초음파 검사는 결절의 발견뿐 아니라 악성과 양성의 감별에 도움을 주며, 초음파유도하 미세침흡인검사에 이용된다. 대표적으로 알려진 악성 결절을 시사하는 초음파 검사 소견으로는 현저한 저에코, 미세석회화, 불규칙한 또는

미세소엽성 경계, 너비보다 큰 키(A/T ratio \geq 1), 및 결절 내 혈류증가 등이 있다^{3,6}. 그러나 초음파 검사는 검사 수행자에 매우 의존적인 주관적인 검사 방법으로써 다양한 관찰자 간의, 또한 관찰자 내의 편차가 존재한다. 이러한 검사자 사이의 또는 검사자 자신의 변이성에 대한 연구는 유방 영상 및 다른 분야에서는 비교적 활발하게 있었다. 그러나 갑상선 진단 영역에서는 갑상선 결절 부피 측정의 관찰자간의 변이에 대한 일부 연구^{7,8} 및 적은 실험군을 대상으로 한 갑상선 양성 결절에서 초음파 판정 일치도 조사에 관한 한 차례의 연구가 있었으나⁹ 악성 및 양성 갑상선 결절의 판정 일치도에 대한 조사 연구는 없었다. 이에 본 연구자는 갑상선을 전공으로 하는 숙련된 영상의학과 의사들을 대상으로 하여 악성 및 양성 갑상선 결절의 모양 및 최종 평가에 대한 관찰자 간과 관찰자 내의 판정 일치도와 평가의 정확성에 대해 알아보고자 한다.

II. 재료 및 방법

1. 대상군의 선택

2001년 1월부터 2007년 10월까지 세브란스 병원에서 갑상선 결절의 미세침흡인검사를 시행 받은 사람 중 갑상선암으로 수술을 받은 환자와 미세침흡인검사로 양성 결절로 확인 받은 후 4년 이상 (47-52개월, 평균 49.52개월) 초음파 추적 검사에서 크기와 모양에 변화가 없거나 거듭된 미세침흡인검사에서도 양성 결절로 판정된 환자를 일차 선택하였다. 이들 중

영상 해상도에 따른 관찰 오차를 줄이기 위하여 두 종류의 초음파기기에서 검사된 영상을 대상으로 하였다. 초음파 기기는 HDI 5000 및 iU 22 (Philips Medical System, Netherlands) 초음파 기기를 이용하였다. 이 영상 중 연구자가 (C.S.H) 임의로 144명의 환자의 204개 결절을 선택하였고, 대상으로 하였다. 사진을 선택한 의사는 판독자에서 제외되었다. 144명의 환자 중 여성은 125명, 남성은 19명 이었으며, 전체 204개의 결절 중 악성은 89예 (89/204, 43.6%), 양성 115예 (115/204, 56.4%)였다. 수술로 확진된 악성 및 양성 결절은 170예, 4년-5년 이상 추적 검사 중인 양성 결절은 34예였다. 1cm 미만의 미세결절은 111예, 1cm 이상 크기의 결절은 93예였다.

2. 초음파 영상의 검토

연구자가 PACS system(GE medical system, USA)에서 각 환자의 초음파 영상 검사에서 회색조의 가로 및 세로 영상과 컬러 도플러 영상에서 2-4장의 대표적인 영상을 선택하여, jpg.file로 변환한 후 마이크로소프트사의 파워포인트 슬라이드 1장에 적절히 배열하여 총 204장의 슬라이드를 만들었다. 전문의 취득 이후 각각 11년, 7년, 5년, 8년 이상 기간의 경력을 보유한, 숙련된 유방 및 갑상선을 전공으로 하는 4명의 영상의학과 전문의(K.E.K., K.J.Y., K.M.J., & S.E.J)들이 개별적으로 컴퓨터 LCD 모니터에서 슬라이드를 검토하였다. 각 결절 및 환자에 대한 임상적 정보는 주어지지 않았으며 관찰자 간의 변이를 조사하기 위하여 표 1과

같은 기술 항목 검토 및 최종 판정을 결절 별로 기록하도록 하였다. 모든 예에서 각 항목에 대해 세부 분류를 시행하였으며 혈류의 경우 컬러도플러 영상이 있었던 123예를 대상으로 하였다. 관찰자 내의 변이성을 조사하기 위해서 4-6주 후 동일한 방법으로 또 한차례 검토하도록 하였다. 각 점검 항목의 기준들에 대해서는 따로 설명하지 않아 기존의 알려진 기준들에 대해 각 검토자들이 얼마나 잘 이해하고 있으며, 서로 일치된 의견을 보이고 있는지를 알아보고자 하였다.

Table1. List of Ultrasonographic Finding Descriptors for Thyroid Nodule and Assessment

Echo pattern	Hyperechoic, Isoechoic, Hypoechoic, Marked hypoechoic/Anechoic
Composition	Solid, Cyst, Mixed-sponge, Mixed-solid \geq 50%, Mixed-solid<50%
Orientation	Parallel, Not-parallel (taller than wide)
Margin	Well-circumscribed, Microlobulated, Irregular/spiculated
Vascularity	Absent, Central, Peripheral, Combined (central & peripheral)
Calcification	Microcalcification Macrocalcification, Both (micro & macro), Eggshell, None
Final	Benign, Probable benign,

3. 통계 분석

검토자 간의, 검토자 내의 일치도를 알아보기 위하여 SAS 시스템 (MAGREE SAS Macro program)을 이용하여 κ 값을 구하였다¹⁰. 총 4명의 관찰자와 6가지 영상 소견 및 최종 판정의 범주 별 κ 값을 측정하기 위해서 Landis와 Koch의 방식을 바탕으로 하였다¹¹. 각 영상 소견에 대해서는 각각 그리고 전체의 κ 값을 구하였고, 경계에 대해서는 각각의 κ 값 외에 “well-circumscribed”와 “irregular/spiculated”와 “microlobulated”를 포함한 “not-circumscribed”로 두 가지로 나눈 것에 대해서도 κ 값을 추가로 구하였다. 혈류의 경우 컬러도플러 영상을 포함한 123예만 대상으로 통계 분석을 시행하였다. 최종 판정 역시 총 4가지 판정에 대하여 κ 값을 구하였고, benign 및 probable benign은 양성으로, low suspicious malignancy 및 suspicious malignancy를 악성으로 하여 두 그룹 간의 κ 값도 추가로 구하였다. κ 값은 1.0이 완전한 일치를 0이 완전 불일치를 의미하며, 0.20이하는 근소한 일치 (slight agreement), 0.21-0.40은 약간의 일치 (fair agreement), 0.41-0.60은 보통의 일치 (moderate agreement), 0.61-0.80은 상당한 일치 (substantial agreement), 0.81-1.00은 거의 완전한 일치 (almost perfect agreement)를 의미한다¹¹.

마지막으로 최종 판정의 benign 및 probable benign은 양성으로, low suspicious malignancy 및 suspicious malignancy를 악성으로 하여 첫 번째로

검토한 데이터를 바탕으로 검토자 개개인의 민감도와 특이도, 양성예측도, 음성예측도 및 정확도를 구하였다.

III. 결과

1. 관찰자 사이와 관찰자 내의 변이성

관찰자 간의 변이성에 대한 κ 값은 표2와 같다. 에코를 제외한 나머지 요소들에 대한 κ 값은 0.41-0.80 사이로 보통에서 상당한 일치율을 보였다. 방위에 대한 평가가 $\kappa=0.61$ 과 0.58로 가장 높은 값을 보였으며, 에코에 대한 평가는 $\kappa=0.34$ 와 0.45로 가장 낮았다. 결절의 조성은 전체 $\kappa=0.59$ 와 0.58로 보통의 일치도를 보이거나 “solid”의 경우 κ 값은 0.71과 0.68로 상당한 정도의 일치도를 나타낸다. 경계의 경우 “well-circumscribed”와 “microlobulated”, “irregular/spiculated”의 세 분류로 나누었을 때에는 κ 값이 차례로 0.53/0.61, 0.35/0.33, 0.23/0.17, 전체 0.42/0.43 이었으나, “well-circumscribed”와 “not-circumscribed”의 두 분류로 나누었을 때에는 κ 값이 각각 0.53과 0.61로 일치도가 매우 증가하였다. 혈류에 대한 판정은 전체 0.46과 0.44로 보통의 일치도를 보였다. 악성 갑상선 결절과 관련이 있다고 알려진 석회화의 판정도 전체 0.58과 0.57로 $\kappa>0.6$ 의 상당한 일치도에 가까운 보통의 일치도를 나타내었다. 마지막으로 최종 판정에서 각각의 κ 값은 benign=0.25/0.44, probable benign=0.66/0.78, low suspicious malignancy=0.46/0.45과 suspicious malignancy=0.49/0.44, 전체 0.54/0.57로

보통의 일치도를 보였으나, 이를 양성과 악성의 두 그룹으로 비교하였을 때에는 κ 값이 각각과 전체 0.72과 0.79로 완벽한 일치에 가까운 상당한 일치를 보인다.

Table2. Inter-observer Variability in Thyroid Nodule Descriptors and Assessment

Category

	Descriptors	1 st κ - value(SE)	2 nd κ - value(SE)
Echo pattern	Hyperechoic	0.13(0.03)	0.42(0.03)
	Isoechoic	0.51(0.03)	0.59(0.03)
	Hypoechoic	0.29(0.03)	0.34(0.03)
	Marked hypo-/Anechoic	0.44(0.03)	0.44(0.03)
	Overall	0.34(0.02)	0.45(0.02)
Composition	Solid	0.71(0.03)	0.68(0.03)
	Cyst	0.75(0.03)	0.66(0.03)
	Mixed-sponge	0.30(0.03)	0.45(0.03)
	Mixed-solid \geq 50%	0.52(0.03)	0.49(0.03)
	Mixed-solid<50%	0.39(0.03)	0.58(0.03)
	Overall	0.59(0.02)	0.58(0.02)
Orientation	Parallel	0.61(0.03)	0.58(0.03)
	Not-parallel	0.61(0.03)	0.58(0.03)

	(taller than wide)		
	Overall	0.61(0.03)	0.58(0.03)
Margin	Well-circumscribed	0.53(0.03)	0.61(0.03)
	Microlobulated	0.35(0.03)	0.33(0.03)
	Irregular/spiculated	0.23(0.03)	0.17(0.03)
	Overall	0.42(0.02)	0.43(0.02)
Margin†	Well-circumscribed	0.53(0.03)	0.61(0.03)
	Not circumscribed	0.53(0.03)	0.61(0.03)
	Overall	0.53(0.03)	0.61(0.03)
Vascularity	Absent	0.42(0.04)	0.41(0.04)
	Central	0.29(0.04)	0.31(0.04)
	Peripheral	0.36(0.04)	0.33(0.04)
	Combined (central & peripheral)	0.67(0.04)	0.63(0.04)
	Overall	0.46(0.02)	0.44(0.02)
Calcification	Microcalcification	0.51(0.03)	0.51(0.03)
	Macrocalcification	0.39(0.03)	0.34(0.03)
	Both (micro & macro)	0.33(0.03)	0.35(0.03)
	Eggshell	0.42(0.03)	0.44(0.03)
	None	0.79(0.03)	0.73(0.03)
	Overall	0.58(0.02)	0.56(0.02)

Final assessment	Benign	0.25(0.03)	0.44(0.03)
	Probable benign	0.66(0.03)	0.78(0.03)
	Low suspicious malignancy	0.46(0.03)	0.45(0.03)
	Suspicious malignancy	0.49(0.03)	0.44(0.03)
	Overall	0.54(0.02)	0.57(0.02)
Final assessment*	Benign	0.72(0.03)	0.79(0.03)
	Malignancy	0.72(0.03)	0.79(0.03)
	Overall	0.72(0.03)	0.79(0.03)

SE: standard error

†: analysis between “well-circumscribed” and “not-circumscribed” (“microlobulated” and “irregular/speculated”) margin

*: analysis between “benign” (“benign” and “probable benign”) and “malignancy” (“low suspicious malignancy” and “suspicious malignancy”)

4명의 검토자 개개인의 전체 κ 값은 표3에 요약하였다. 대부분의 기술어 및 판정의 κ 값이 0.61이상의 상당한 일치를 보여 주었다. 경계의 경우 “well-circumscribed”와 “microlobulated” 또는 “irregular/speculated”로 나누었을 때보다 “we---circumscribed”와 “not-circumscribed”로 나뉘어졌을 때 κ 값이 증가하였다. 특히 final assessment는 두 group으로 나누었을 때에 거의 완벽한 일치도를 보였다.

Table3. Intra-observer Variability in Thyroid Nodule Descriptors and Assessment

Category

Descriptors	κ-value(SE)			
	Observer 1	Observer 2	Observer 3	Observer 4
Echo pattern	0.65(0.05)	0.74(0.05)	0.47(0.05)	0.69(0.05)
Composition	0.69(0.05)	0.83(0.06)	0.78(0.05)	0.76(0.05)
Orientation	0.74(0.07)	0.88(0.07)	0.75(0.07)	0.79(0.07)
Margin	0.60(0.07)	0.84(0.05)	0.67(0.05)	0.63(0.07)
Margin†	0.64(0.07)	0.93(0.07)	0.77(0.07)	0.66(0.07)
Vascularity	0.64(0.07)	0.90(0.06)	0.76(0.06)	0.72(0.06)
Calcification	0.74(0.05)	0.87(0.04)	0.66(0.04)	0.77(0.04)
Final assessment	0.78(0.06)	0.87(0.06)	0.51(0.05)	0.76(0.05)
Final assessment*	0.88(0.07)	0.95(0.07)	0.67(0.07)	0.91(0.07)

SE: standard error

†: analysis between “well-circumscribed” and “not-circumscribed” (“microlobulated” and “irregular/speculated”) margin

*: two group analysis between benign (benign & probable benign) and malignancy (low suspicious malignancy & suspicious malignancy)

2. 관찰자 개인 판정의 정확성

각 관찰자 별로 최종 판정의 benign군과 probable benign군을 양성으로

low suspicious malignancy군과 suspicious malignancy군을 악성으로 하여 첫 번째 시행한 검토키록을 바탕으로 각각의 민감도, 특이도, 양성예측도, 음성예측도 및 정확도를 구하였다(표4). 모든 검토자에서 80%이상의 정확도를 보였으며, 음성예측도는 85%이상으로 매우 높았다.

Table4. Individual Data Correlation 1st Final Assessment with Pathology

Observer	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy (%)
1	83/89 (93.3)	88/115 (76.5)	83/110 (75.5)	88/94 (93.6)	171/204 (83.8)
2	75/89 (84.3)	89/115 (77.4)	75/101 (74.3)	89/103 (86.4)	164/204 (80.4)
3	74/89 (83.1)	100/115 (87.0)	74/89 (83.1)	100/115 (87.0)	174/204 (85.3)
4	82/89 (92.1)	85/115 (73.9)	82/112 (73.2)	85/92 (92.4)	167/204 (81.9)
Overall	314/356 (88.2)	362/460 (78.7)	314/412 (76.2)	362/404 (89.6)	676/816 (82.8)

PPV: positive predictive value, NPV: negative predictive value

IV. 고찰

갑상선 결절을 비교적 흔한 내과적 질환으로 약 19-67%의 유병률을 보인다¹. 갑상선 질환에서 초음파 검사는 만져지지 않는 갑상선 결절을

정확하게 탐지하며, 결절의 크기 및 모양, 갑상선 선종의 부피 등을 평가할 수 있는 중요한 진단 기기이다. 또한 고해상도 초음파 기기의 등장 이후 과거 결절의 유무 또는 양성인지 고형성 종괴인지의 여부만 판단하던 것에서 나아가 양성 결절과 악성 결절의 감별에 관한 연구 결과들이 발표되었다. 그 결과 종괴 내 미세석회화, 현저한 저에코, 불규칙적인 경계, 가로 너비 보다 긴 세로 너비, 혈류의 모양, 일부 거대석회화 등이 갑상선암과 관련이 있는 소견이라 알려 졌다.^{3, 4, 6, 12, 13}. 이러한 소견들은 갑상선 결절 내에 개별적으로 혹은 동시에 관찰되며, 초음파 검사를 시행하는 의사는 각자의 경험 및 지식에 바탕을 두고 결절에 대한 평가를 내린다. 그러나 초음파 검사는 검사자에 매우 의존적인 주관적인 검사로 검사를 시행하는 의사에 따라 그 정확성에 차이가 있다. 이러한 점에 기초하여 유방 및 다른 영역에서의 초음파 검사에 관한 관찰자 간의 검사의 정확성에 대한 연구 및 일치도 연구가 있어 왔으나^{14, 15} 갑상선 결절에 대한 연구는 부피 측정을 제외하고는^{7, 16} 거의 전무하다. 따라서 본 연구에서는 유방과 갑상선을 전공으로 한 숙련된 영상의학과 전문의들의 갑상선 결절의 초음파 영상 소견의 기술 및 판정의 일치도 및 정확도를 알아보고자 하였다.

본 연구 결과 숙련된 영상의학과 의사들간, 의사 본인의 결절 기술어 선택 및 판정 일치도는 대부분에 있어 보통 이상과 상당한 정도의 일치도를 보였다 (표2 참조). 가장 낮은 일치도를 보인 것은 결절 에코의 판정으로 전체 κ 값은 0.34였다. 이는 아마도 갑상선 결절의 에코는 갑상선

실질 에코를 기준으로 평가하는데, 미만성 갑상선 질환이 있는 경우 갑상선 실질 에코에 변화가 있으므로 인해, 혹은 결절 내 석회에 의한 영향 등으로 평가에 어려움이 있었을 것으로 여겨진다 또한 갑상선 결절 자체가 다양한 구성 요소를 포함함으로써 인해 불균질한 에코를 보이는 경우에도 한 가지 에코로 평가하기 어렵다(그림1, 2).

Fig1. A 46-year-old woman with a papillary carcinoma. US shows a heterogeneous echoic nodule in the upper pole of the right thyroid lobe. Two observers described hypoechoic nodule, but the others chose marked hypoechoic nodule. The margin, orientation and final assessment were assessed as well-circumscribed, parallel and suspicious malignancy by all observers.

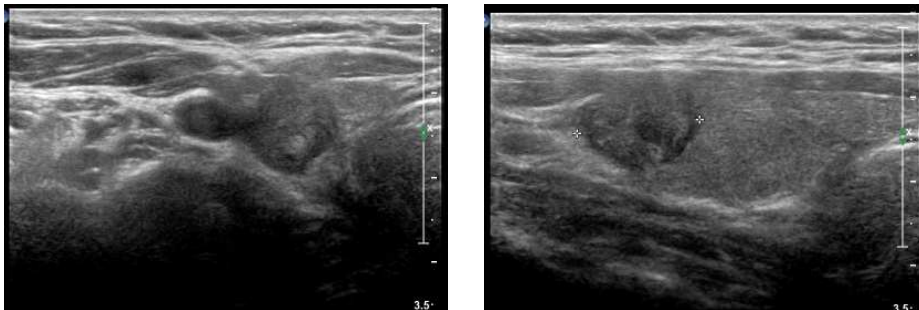
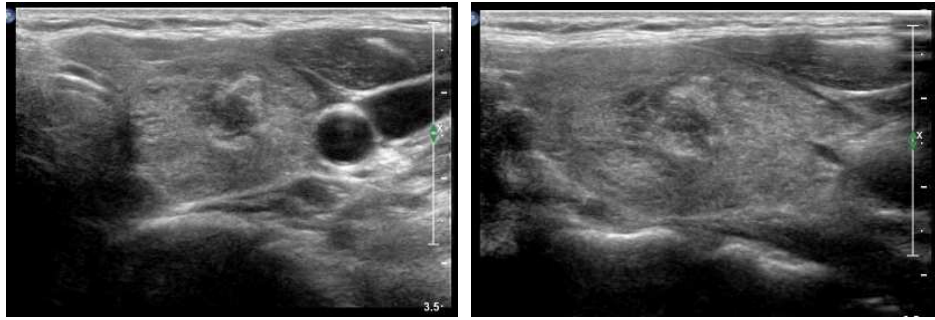
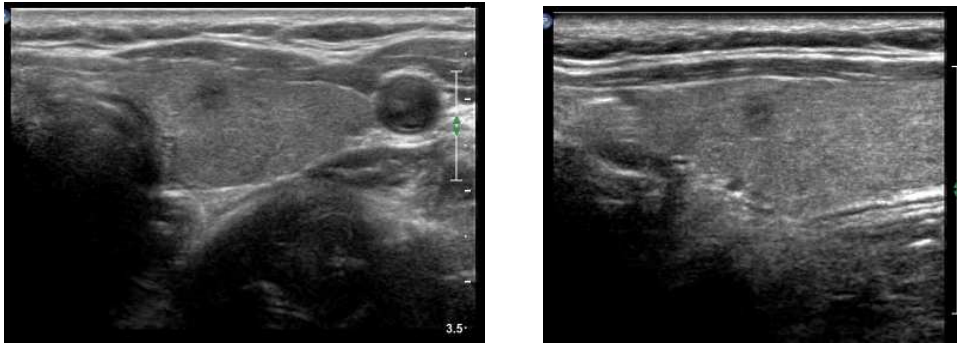


Fig2. A 18-year-old woman with a papillary carcinoma. US reveals a heterogeneously echoic nodule with calcification in the mid pole of the left thyroid lobe. Two observers described hypoechoic nodule, one chose marked hypoechoic, and another selected hyperechoic nodule. Other descriptors are all agreed by four observers.



경계의 경우 “well-circumscribed”와 “microlobulated”, “irregular/spiculated”의 세 분류로 나누었을 때에는 κ 값이 차례로 0.53/0.61, 0.35/0.33, 0.23/0.17, 전체 0.42/0.43로 약간 혹은 보통의 일치도를 보였으나 “well-circumscribed”와 “not-circumscribed”의 두 분류로 나누었을 때에는 κ 값이 각각 0.53과 0.61로 일치도가 증가하였다. 임상적으로 “microlobulated”나 “irregular” 경우 모두 의심스러운 초음파 소견이므로 “well-circumscribed”와 “not-circumscribed”의 두 군으로의 일치도가 높아 임상적인 어려움은 없을 것으로 생각된다. (그림2,3)

Fig3. A 55-year-old woman with multifocal papillary carcinomas. A small hypoechoic nodule with not-parallel orientation is visible in the upper pole of the left thyroid lobe. Two observers described microlobulated margin, but the others chose irregular or spiculated margin. However both of them are not-circumscribed margin.



그러나 본 연구에서 흥미로운 결과는 각 판정 항목에서 가장 악성도와 관련이 있다고 알려진 기술어, 즉 “marked hypoechogenicity”, “solid”, “microcalcification”의 κ 값이 평균보다 높거나 가장 높았다는 것이다 (표2 참고). 이러한 점은 숙련된 영상의학과 전문의는 갑상선 결절의 평가에 있어 중요하다고 알려진 기준을 잘 숙지하고 있으며, 그 기준에 특별히 주의를 기울여 검사를 수행하고 있음을 시사한다고 볼 수 있겠다. 또한 최종 판정을 단순히 양성과 악성의 두 그룹으로 나누어 비교하였을 때의 κ 값이 전체 κ 값 중 가장 높은 수준의 일치도($\kappa=0.72$)를 보인 것은 최종 판정의 차이가 각각의 기술어의 선택의 차이보다 적다는 것을 의미한다고 할 수 있겠다. 또한 그 판정의 정확성은 80% 이상의 높은 수준을 보였다 (표4).

관찰자 내의 일치도는 관찰자 간의 일치도 보다 높은 κ 값을 보였다. (표3) 이는 각 영상의학과 전문의가 고유의 일관된 판단 기준을 가지고 있음을 시사한다고 볼 수 있겠다. 또한 개개인의 판정 정확도는 병리 결과와 비교하였을 때의 민감도, 음성 예측도, 정확도가 80% 이상으로

높은 수준을 보였다(표4). 또한 전체 양성 예측도는 76.2%로 이전 Nóbrega 등의 연구에서 발표한 고위험군의 양성 예측도 75% 보다 약간 더 높은 수치로¹⁷ 연구에 참여한 영상의학과 전문의들이 적절한 판정을 시행하고 있다는 것을 시사한다.

V. 결론

결론적으로 숙련된 영상의학과 전문의들은 각각 갑상선 결절 평가에 있어 갑상선암과 관련된 영상 소견을 잘 숙지하고 있으며, 그 평가에 있어서로 간, 그리고 본인 내에서 비교적 높은 수준의 일치도를 보였다. 또한 병리 결과와 결절의 판정을 비교하였을 때에도 높은 수준의 정확도를 보였다.

참고문헌

1. Tan GH, Gharib H. Thyroid incidentalomas: management approaches to nonpalpable nodules discovered incidentally on thyroid imaging. *Annals of Internal Medicine* 1997;126:226-31.
2. Hegedüs L. Clinical practice. The thyroid nodule. *New England Journal of Medicine* 2004;351:1764-71.
3. Kim EK, Park CS, Chung WY, Oh KK, Kim DI, Lee JT, et al. New sonographic criteria for recommending fine-needle aspiration biopsy of nonpalpable solid nodules of the thyroid. *AJR, Am J Roentgenol* 2002;178:687-91.
4. Cappelli C, Pirola I, Cumetti D, Micheletti L, Tironi A, Gandossi E, et al. Is the anteroposterior and transverse diameter ratio of nonpalpable thyroid nodules a sonographic criteria for recommending fine-needle aspiration cytology? *Clinical Endocrinology* 2005;63:689-93.
5. Cappelli C, Castellano M, Pirola I, Cumetti D, Agosti B, Gandossi E, et al. The predictive value of ultrasound findings in the management of thyroid nodules. *QJM* 2007;100:29-35.
6. Papini E, Guglielmi R, Bianchini A, Crescenzi A, Taccogna S, Nardi F, et al. Risk of malignancy in nonpalpable thyroid nodules: predictive value of ultrasound and color-Doppler features. *Journal of Clinical Endocrinology and Metabolism* 2002;87:1941-6.

7. Brauer VF, Eder P, Miehle K, Wiesner TD, Hasenclever H, Paschke R. Interobserver variation for ultrasound determination of thyroid nodule volumes. *Thyroid* 2005;15:1169-75.
8. Slapa RZ, Slowinska-Srzednicka J, Szopinski KT, Jakubowski W. Gray-scale three-dimensional sonography of thyroid nodules: feasibility of the method and preliminary studies. *European Radiology* 2006;16:428-36.
9. Wienke JR, Chong WK, Fielding JR, Zou KH, Mittelstaedt CA. Sonographic features of benign thyroid nodules: interobserver reliability and overlap with malignancy. *J Ultrasound Med* 2003;22:1027-31.
10. Fleiss JL. *Statistical methods for rates and proportions*. 2nd ed. New York: John Wiley & Sons Inc; 1981.
11. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
12. Iannuccilli JD, Cronan JJ, Monchik JM. Risk for malignancy of thyroid nodules as assessed by sonographic criteria: the need for biopsy. *J Ultrasound Med* 2004;23:1455-64.
13. Kovacevic DO, Kovacevic O, Skurla MS. Sonographic diagnosis of thyroid nodules: correlation with the results of sonographically guided fine-needle aspiration biopsy. *J Clin Ultrasound* 2007;35:63-7.
14. Lazarus E, Mainiero MB, Schepps B, Koelliker SL, Livingston LS. BI-RADS lexicon for US and mammography: interobserver variability and positive

predictive value. *Radiology* 2006;239:385-91.

15. Lee HJ, Kim EK, Kim MJ, Youk JH, Lee JY, Kang DR, et al. Observer variability of Breast Imaging Reporting and Data System (BI-RADS) for breast ultrasound. *European Journal of Radiology* 2008;65:293-8

16. Salih Deveci M, Deveci G, Livolsi VA, Gupta PK, Baloch ZW. Concordance between thyroid nodule sizes measured by ultrasound and gross pathology examination: Effect on patient management. *Diagn Cytopathol* 2007;35:579-83.

17. Nóbrega LH, Paiva FJ, Nóbrega ML, Mello LE, Fonseca HA, Costa SO, et al. Predicting malignant involvement in a thyroid nodule: role of ultrasonography. *Endocrine Practice* 2007;13:219-24.

Abstract

Inter-observer Variation for US Assessment of Thyroid Nodule

Choi Seon Hyeong

Department of Medicine
The Graduate School, Yonsei University

(Directed by Professor Kim Eun-Kyung)

Thyroid nodules are common medical problem and their first diagnostic tool is US equipment. However ultrasonographic examination is very dependant on operators so there are observer variations. In this study we evaluated inter- and intra-observer variations for US assessment of thyroid nodule description and diagnosis by four radiologists, additionally estimated the diagnostic accuracy.

A total of 204 thyroid nodules of 144 patients were reviewed; there were 89 cases of benign and 115 cases of malignancy. Four radiologists, who have more than 5 year of experience, had reviewed independently twice with 4-6 weeks intervals. The echogenicity, composition, margin, orientation, calcification, vascularity and final assessment were evaluated. Inter- and intra-observer variations were assessed with Cohen's kappa statistics and accuracy was calculated.

For inter-observer variations, the echogenicity showed slight agreement ($\kappa=0.34$), fair agreements ($\kappa=0.59$, $\kappa=0.42$, $\kappa=0.58$, and $\kappa=0.54$, respectively) were found in the composition, margin, calcification and final assessment, and

substantial agreements ($\kappa=0.61$ and $\kappa=0.64$, respectively) were noted in the orientation and vascularity. For intra-observer variability, almost of them showed substantial agreement more than $\kappa>0.61$. The overall sensitivity, specificity, positive predictive value, negative predictive value, and accuracy of four radiologists were as followed; 88.2%, 78.7%, 76.2%, 89.6%, and 82.8%.

In conclusion, experienced radiologists showed more than the moderate degree of agreement in US assessment for thyroid nodule, and their final assessment were quietly accurate.

Key Words: thyroid nodule, sonography, interobserver variation, intraobserver variation