

갑상선 초음파: 교육에 따른 검사자간 일치도와 진단 정확도의 변화

문희정 · 김은경 · 박정선²
곽진영

¹연세대학교 의과대학 영상의학과

²한양대학교 의과대학 영상의학과

J Korean Soc Ultrasound Med

2011;30:23-28

Received October 30, 2010; Revised
January 20, 2011; Accepted January 31,
2011.

이 연구는 대한초음파의학회에서 지원한
2009년 필립스 장학기금으로 진행되었다.
This study was supported by 2009
Philips Scholarship Fund through Korean
Society of Ultrasound in Medicine.

Address for reprints :

Jin Young Kwak, MD, Department of
Radiology, Research Institute of
Radiological Science, Yonsei University
College of Medicine, 250 Seongsanno,
Seodaemun-gu, Seoul 120-752, Korea.
Tel. 82-2-2228-7400
Fax: 82-2-393-3035
E-mail: docjin@yuhs.ac

Thyroid Ultrasound: Change of Inter-observer Variability and Diagnostic Performance after Training

Hee Jung Moon, MD, Eun-Kyung Kim, MD, Jeong Seon Park, MD²,
Jin Young Kwak, MD

¹Department of Radiology, Research Institute of Radiological Science, Yonsei University
College of Medicine

²Department of Radiology, Hanyang University Medical Center

Purpose: To investigate and compare inter-observer variability and diagnostic performance of thyroid ultrasound (US) between a faculty member and observing residents.

Materials and Methods: From October 2007 to June 2009, 18 residents underwent training in thyroid US section. Group 1 included 8 residents that were trained for the first time and group 2 included 10 residents that were trained for the second time. US features of nodules were recoded according to the composition, echogenicity, margin, calcifications, shape, and final assessment by a faculty member and residents, respectively. Following a discussion, a faculty member performed fine needle aspiration. Then, the inter-observer variability and diagnostic performance between a faculty member and residents were investigated and compared for US.

Results: In group 1, agreement for composition in resident 1, calcification for residents 5 and 6, and shape for resident 4 were slight, moderate, moderate, and moderate, respectively. In group 2, agreement for composition in residents 1 and 10 were moderate. Substantial or greater agreement was observed more frequently in group 2 than 1. The diagnostic performances for both the faculty and residents were high and not statistically different.

Conclusion: Agreement for US features between a faculty and residents as well as diagnostic performance were high. Moreover, diagnostic performance of residents that underwent training a second time of training was higher than residents that underwent training only once.

Key words : Thyroid; Ultrasound (US); Thyroid, biopsy; Inter-observer variability

서론

갑상선 결절은 매우 흔한 질환으로 인구의 4-7%가 결절을 가지고 있다 [1, 2]. 초음파는 갑상선 결절의 발견뿐 아니라 양성, 악성 결절의 감별에 매우 중요한 역할을 한다 [3-7]. 악성을 높게 예측할 수 있는 초음파 소견은 주변 근육보다 현저히

낮은 저에코 (marked hypoechogenicity), 소분엽성 혹은 불규칙한 변연부 (microlobulated or irregular margins), 전후 장경이 넓이 보다 큰 모양 (taller than wide shape), 미세 석회화 (microcalcifications)로 알려져 있으며, 각각의 소견 보다는, 이 소견들의 조합이 더 높은 진단 정확도를 보인다 [4-6]. 2009년 개정된 American Thyroid Association guidelines에 따르면 저에코 (hypoechogenicity), 소분엽성 혹은 불규칙한

변연 (microlobulated or irregular margins), 전후 장경이 넓이 보다 큰 모양 (taller than wide shape), 미세 석회화 (microcalcifications), 결절내 혈류량 증가를 악성이 의심되는 소견으로 보았다 [8]. 갑상선 암의 고위험군 환자에 있어 위의 소견들이 보이면 5 mm 이상의 작은 갑상선 결절에서도 세침 흡입생검을 권한다 [8]. 여러 초음파 소견들의 조합이 진단 정확도가 높기는 하나, 초음파는 주관적이며 시술자 의존적인 검사이다. 따라서 검사자간 일치도가 문제가 될 수 있다. 갑상선 초음파에 경험이 많은 전문가들간에 갑상선 결절에 대한 초음파 소견의 검사자간 일치도를 보는 이전 연구들이 있었고, 그 결과 높은 일치도를 보였다 [9, 10]. 그러나 경험이 없는 초심자에서의 교육 전후의 검사 일치도를 비교해 보는 연구는 없었다.

따라서 본 연구는 갑상선 초음파 파트에서 교련을 받은 전공의를 대상으로 교육 정도에 따라 갑상선 전문의와의 검사자간 일치도와 진단 정확도의 변화를 알아 보고자 하였다. 또한 교육을 통해 일치도가 향상 되는지 알아 보았다.

대상 및 방법

이 연구는 후향적 연구로 임상시험심사위원회의 승인을 받았다. 2007년 10월부터 2009년 6월까지 18명의 전공의가 갑상선 초음파 파트에서 교육을 받았다. 본 기관에서는 전공의 수련과정 중 갑상선 초음파는 저년차 (1, 2년차)와 고년차 (3, 4년차) 때 한 달씩 총 두 번, 두 달 동안 교육을 받는다. 한 달 동안 전문의와 일대일로 실시간 초음파를 보면서 경험을 공유하는 시기는 스케줄마다 다르며 가능한 한 열 명 이상의 환자의 증례를 교육할 기회를 가지고자 하였다. 다섯 명의 전공의는 연구 기간 동안 처음 한 달의 교육을 받았으며, 일곱 명의 전공의는 연구 기간 이전에 한 번을 받고 연구 기간 동안 두 번째 교육을 받아 총 두 달의 교육을 받았다. 세 명의 전공의는 연구 기간 동안 각각 다른 연도에 두 번 교육을 받아 총 두 달의 교육을 받았다. 따라서 열 다섯 명의 전공의가 연구 기간 동안 열 여덟 번의 교육을 받았다. 한 번 교육을 받은 전공의를 그룹 1로 분류하였으며 총 여덟 명의 전공의로 구성 되었다. 두 번 교육을 받은 전공의를 그룹 2로 분류하였으며 총 열 명의 전공의로 구성 되었다. 연구 기간 동안 총 664 갑상선 결절에 대해 전문의가 초음파 시행 직후 세침흡입생검을 시행하였다. 268 결절은 그룹 1 전공의가 같이 참관하였으며 396 결절은 그룹 2 전공의가 참관하였다. 세침흡입생검의 결과가 불충분 검체로 나온 결절은 99개 였다 (15%). 36개의 결절은 세침흡입생검상 유두암 의심으로 진단 되었고, 이 중 30개의 결절이 수술을 받았다. 불충분 검체로 나온 결절과 유두암 의심 결절 중 수술을 시행하지 않은 결절을 제외한 559 결절이 이 연구에 포함 되었다. 연구 기간 중 세침흡입생검상 소포신생물 (follicular neoplasm)의 결과를 얻은 결절은 없었다. 전체 559 결절중

196 결절에 대해 수술을 시행했고 363 결절에 대해서는 세침 흡입생검만 시행했다. 총 559 결절 중, 221 결절 (39.5%, 221 of 559)은 그룹 1 전공의가, 338 결절 (62.5%, 338 of 559)은 그룹 2 전공의가 초음파와 세침흡입생검을 참관하였다.

초음파 기기는 5- to 12-MHz 탐침이 부착된 iU22 (Philips Medical Systems, Bothell, WA)를 이용 하였다. 모든 초음파는 갑상선 전문의가 시행하였고 실시간 초음파를 전공의가 같이 참관하였다. 초음파 후 초음파 소견을 전공의와 전문의가 각각 기입하고, 이 후 사진을 보면서 다시 검토를 하였다. 판독 결과지는 구성성분, 에코, 변연부, 석회화, 모양, 최종 결과에 따라 Table 1과 같이 분류하였다. 악성 의심 소견은 현저히 낮은 저에코 (marked hypoechogenicity), 소분엽성 혹은 불규칙한 변연 (microlobulated or irregular margins), 전후 장경이 넓이 보다 큰 모양 (taller than wide shape), 미세 석회화 (microcalcifications)로 정하였으며, 이중 하나 이상의 소견이 보이면 악성 의심 (suspicious)로, 악성 의심 소견이 하나도 없는 경우 양성 (probably benign)으로 분류 하였다 [4, 5]. 악성 의심 결절이 있거나 악성 의심 결절이 없는 경우 양성 결절 중 가장 크기가 큰 결절에 대해 세침흡입생검을 전문의가 시행 하였다. 23 게이지 바늘과 2 cc 시린지를 이용하여 한 결절당 최소 2번의 세침흡입생검을 시행하였다.

전공의와 갑상선 전문의가 기록한 각각의 초음파 소견을 Cohen의 Kappa 분석을 이용하여 k 값이 0.2 이하는 불량 (slight agreement), 0.21 - 0.40는 보통 (fair agreement), 0.41

Table 1. Classifications of Ultrasound Features of Thyroid Nodules

Composition
Solid
Cystic portion \leq 50%
Cystic portion $>$ 50%
Echogenicity
Hyperechogenicity
Isoechogenicity
Hypoechogenicity
Marked hypoechogenicity
Margin
Well defined (n = 1036)
Microlobulated (n = 395)
Irregular (n = 62)
Calcifications
Microcalcification
Macrocalcification
Mixed calcifications
Egg shell calcification
No calcification
Shape
Wider than tall
Taller than wide
US Final Assessment
Probably benign
Suspicious

- 0.60는 중중도 (moderate agreement), 0.61 - 0.80은 우수 (substantial agreement), 0.81 - 1.00은 아주 우수 (almost perfect agreement)로 분류하였다 [11]. 악성과 양성을 판단하는 기준은 세침흡입생검의 결과와 수술 결과로 하였다. 갑상선암에 대한 진단 정확도는 receiver operating characteristic (ROC) curve 분석을 이용하였고, under the ROC curve (Az 값)를 이용하여 비교하였다. 분석은 SPSS (version 18.0 for Windows; SPSS, Chicago, IL)를 이용하였다.

결 과

559 결절 중 수술을 통해 176 결절은 악성, 20 결절은 양성으로 진단 되었다. 372 결절은 세침흡입생검 상 양성, 11 결절은 악성의 결과를 얻었다. 따라서, 187 결절은 악성으로 372 결절을 양성으로 분류 되었다. 187 악성중 184개는 유두암, 2개는 여포성 유두암, 1개는 수질암 이었다. 그룹 1 전공의가 참관한 221 결절 중 145개는 양성 그리고 76개는 악성이었다.

그룹 2 전공의가 참관한 338 결절 중 227 양성이고 111개는 악성 이었다.

8명의 그룹 1 전공의와 갑상선 전문의간 초음파 소견의 검사자간 일치도는 Table 2와 같다. 전공의 1은 구성 성분에서 0.121의 낮은 κ 값을 보였다 (Fig. 1). 전공의 5와 6은 석회화에서 (Fig. 2), 전공의 4는 모양에서 중중도의 일치도를 보였다. 그 외는 모든 초음파 소견은 우수 이상의 일치도를 보였다. 진단 정확도를 나타내는 Az 값은 0.608 - 1.000으로 다양했으며 전문의와 비교 시 통계학적으로 큰 차이를 보이지 않았다. Az 값이 전공의 5와 전문의 모두에서 1 이었으며 이때 검사한 결절 수는 8개로 적었다.

그룹 2 전공의의 검사자간 일치도와 진단 정확도는 Table 3과 같다. 전공의 1과 10의 구성 성분에 대한 κ 값이 중중도의 일치도를 보였다. 그외 모든 초음파 소견에서 우수 이상의 일치도를 보였으며 Az 값도 전문의와 통계학적인 차이가 없었다.

그룹 1과 그룹 2, 두 그룹간 진단 일치도는 group 1에서 중중도 이하의 일치도를 보이는 경우가 그룹 2 보다 더 많았다.

Table 2. Interobserver Variability and Diagnostic Performances between Residents in Group 1 and a Faculty Radiologist

	N	Com	Echo	Margin	Calcification	Shape	Vascularity	Final	Az of Resident (95% CI)	Az of Faculty (95% CI)	P value
R1	25	0.121	0.667	0.514	0.768	0.839	0.747	0.762	0.812 (0.607 - 0.938)	0.906 (0.721 - 0.983)	0.375
R2	36	0.772	0.728	0.648	0.832	0.609	0.957	0.809	0.736 (0.563 - 0.869)	0.754 (0.582 - 0.881)	0.800
R3	40	0.953	0.913	0.778	0.649	0.939	0.811	0.780	0.811 (0.655 - 0.917)	0.868 (0.723 - 0.953)	0.449
R4	28	0.872	0.881	0.741	0.883	0.417	1.000	0.784	0.862 (0.678 - 0.961)	0.828 (0.639 - 0.943)	0.601
R5	8	0.619	1.000	1.000	0.556	0.600	1.000	1.000	1.000 (0.629 - 1.000)	1.000 (0.629 - 1.000)	1.000
R6	34	0.858	0.815	0.708	0.458	0.751	0.955	0.765	0.802 (0.630 - 0.918)	0.913 (0.765 - 0.981)	0.231
R7	27	1.000	0.768	0.660	0.817	0.924	0.887	0.702	0.792 (0.593 - 0.922)	0.933 (0.767 - 0.991)	0.136
R8	23	0.736	0.675	0.693	0.703	0.587	0.862	0.649	0.608 (0.385 - 0.802)	0.771 (0.550 - 0.917)	0.129

N; number of thyroid nodules

R; resident

Az; under the receiver operating characteristic curve

CI; confidence interval

Final; final assessment on US

Table 3. Interobserver Variability and Diagnostic Performances between Residents in Group 2 and a Faculty Radiologist

	N	Com	Echo	Margin	Calcification	Shape	Vascularity	Final	Az of Resident (95% CI)	Az of Faculty (95% CI)	P value
R1	38	0.515	0.718	0.623	0.711	0.653	0.808	0.838	0.912 (0.773 - 0.978)	0.846 (0.692 to 0.942)	0.219
R2	42	0.836	0.953	0.753	0.762	0.877	1.000	0.668	0.838 (0.692 - 0.933)	0.794 (0.641 - 0.903)	0.710
R3	11	1.000	0.825	0.828	0.825	0.621	1.000	1.000	0.732 (0.395 - 0.939)	0.732 (0.395 - 0.939)	1.000
R4	48	0.655	0.862	0.618	0.953	0.684	0.964	0.958	0.859 (0.729 - 0.942)	0.875 (0.747 - 0.952)	0.483
R5	52	1.000	1.000	0.797	0.964	0.902	0.915	0.923	0.808(0.675 - 0.904)	0.833 (0.704 - 0.9220)	0.761
R6	47	0.745	0.921	0.809	0.928	0.682	0.936	0.912	0.830 (0.692 - 0.923)	0.802 (0.660 - 0.904)	0.623
R7	18	0.811	0.789	0.647	0.811	0.824	0.903	0.620	0.893 (0.657 - 0.984)	0.839 (0.593 - 0.965)	0.730
R8	26	0.764	0.726	0.653	0.786	0.923	0.938	0.830	0.798 (0.595 - 0.928)	0.875 (0.686 - 0.970)	0.397
R9	37	0.921	0.926	0.780	0.838	0.916	0.951	0.780	0.705 (0.532 - 0.843)	0.813 (0.651 - 0.921)	0.094
R10	19	0.441	0.882	0.658	0.744	0.894	0.823	1.000	0.792 (0.547 - 0.939)	0.792 (0.547 - 0.939)	1.000

N; number of thyroid nodules

R; resident

Az; under the receiver operating characteristic curve

CI; confidence interval, Final; final assessment on US

고 찰

갑상선 질환에 있어 초음파는 쉽게 이용할 수 있고, 정확하고, 방사선 피폭이 없는 검사로 널리 사용 된다. 초음파는 갑상선 결절의 발견, 양성과 악성의 감별, 세침흡입생검 유도등에 유용하게 이용할 수 있다 [3-7]. 현저히 낮은 저에코 (marked hypoechogenicity), 소분엽성 혹은 불규칙한 변연 (microlobulated or irregular margins), 전후 장경이 넓어 보다 큰 모양 (taller than wide shape), 미세 석회화 (microcalcifications) 소견은 여러 연구를 통해 갑상선 암을 진단하는데 유용한 소견으로 확인 되었고, 특히 이들의 조합이 높은 진단 정확도를 보였다 [3-7]. 이런 소견들이 진단에 유

용하게 이용되려면 소견들의 진단 정확도도 높아야 하지만 검사의 재현성이 중요하다. 초음파 검사는 매우 주관적이고 시술자에 영향을 많이 받는 검사이기 때문이다. 갑상선 결절의 초음파 소견의 일치도를 보는 연구가 경험 많은 영상의학전문의를 기준으로 진행되었고 악성이 의심되는 초음파 소견뿐만 아니라 그 외의 다른 초음파 소견들의 검사자간 일치도는 높았다 [9]. 경험 없는 의사에서 초음파의 진단적 가치를 높이기 위해서는 경험 많은 의사와의 경험의 공유가 중요하고 교육에 의한 초음파의 진단정확도를 높일 수 있다 [12]. 따라서 이 연구에서는 경험이 적은 전공의를 대상으로 교육의 노출 횟수에 따라 갑상선 결절의 초음파 소견에 대한 검사 일치도와 갑상선 암의 진단정확도에 대한 연구를 진행하였다.

그룹 1은 처음 갑상선 초음파 파트에서 공부를 한 여덟 명의

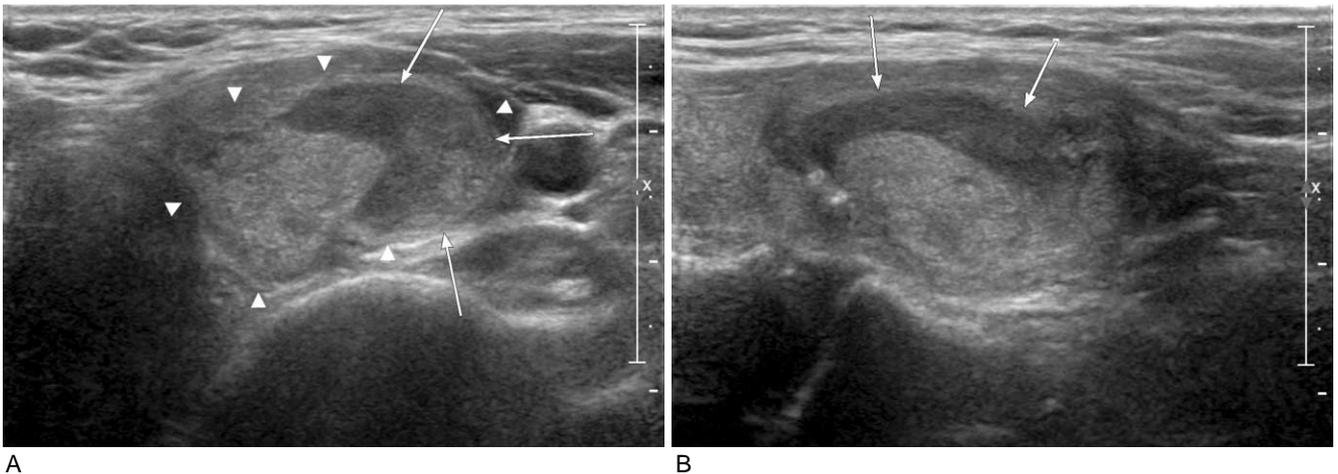


Fig. 1. A 55-year-old woman presented palpable neck mass. About 2.7 cm size mass (arrowheads) in left thyroid gland showed isoechoogenicity, well defined margin, wider than tall shape, and no calcifications on US (A; transverse image and B; longitudinal image). The mass was classified as mass with cystic portion \leq 50 % by a resident because of the hypoechoic portion (arrows) but as solid by a faculty. Adenomatous hyperplasia was diagnosed by fine needle aspiration.

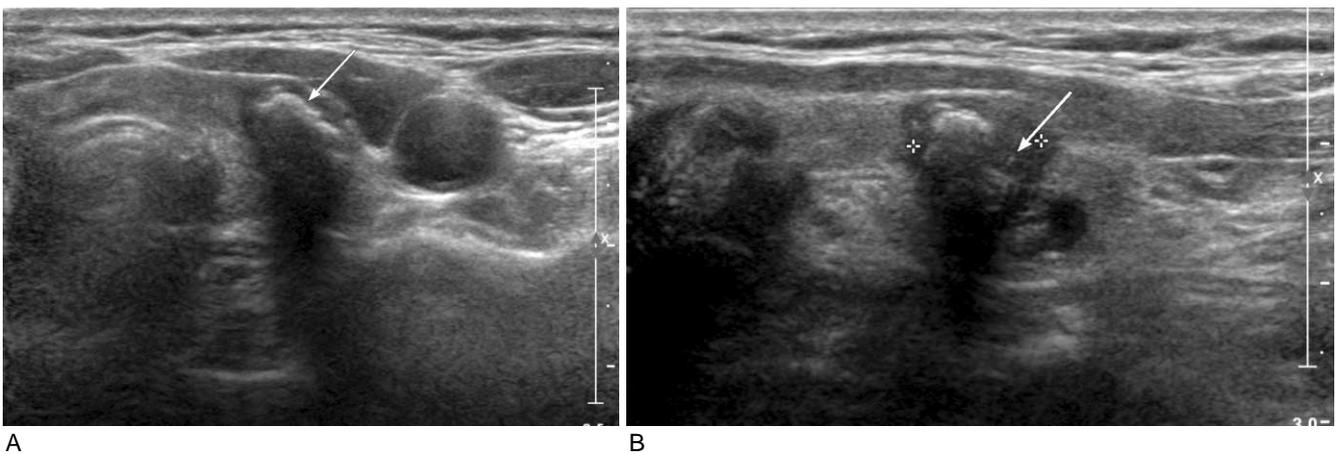


Fig. 2. A 65-year-old woman underwent routine checkup. The 1.2 cm size mass in left thyroid gland showed solid, hypoechoogenicity, microlobulated margin, and taller than wide shape on US (A; transverse image and B; longitudinal image). Calcifications within the mass were classified as macrocalcifications by a resident but as mixed calcifications by a faculty (both of macro- and microcalcifications (arrow)). Papillary thyroid carcinoma was diagnosed by fine needle aspiration and surgery.

전공의로 구성 되었다. 대부분의 초음파 소견은 우수 이상의 높은 일치도를 보였다. 그러나 전공의 1은 구성 성분에서 k 값이 0.121로 불량한 일치도를 보였다. 이는 결절의 내부에 작은 낭성 부분을 찾지 못하거나, 50% 기준에 대해 검사자간 차이가 있어서 일 것으로 생각된다. 전공의 1은 변연부에서 중증도의 일치도를 보여 다른 전공의 보다 낮은 일치도를 보였다. 이는 경계가 잘 지워지는 결절과 그렇지 않은 결절은 잘 구분을 하나, 소엽성과 불규칙한 변연부의 분류가 원인이 되었을 것으로 생각된다. 전공의 5, 6은 석회화에 대해 중증도의 일치도를 보여 다른 소견에 비해 낮은 일치도를 보였다. 거대 석회화 (macrocalcification)와 계란 껍질 모양 석회화 (egg shell calcification)의 분류, 미세 석회화 (microcalcifications)와 혼합 석회화 (mixed calcifications)의 분류는 다른 의견을 많을 수 있고, 이 때문에 낮은 일치도를 보였을 것으로 생각된다.

그룹 2는 갑상선 초음파 파트를 두 번째 수련하는 열 명의 전공의로 구성 되었다. 전공의 1과 10이 구성 성분에서 다른 소견보다 낮은 중증도의 일치도를 보였다. 이 역시 앞서 설명한 그룹 1에서만 마찬가지로 이는 결절의 내부에 작은 부분이 낭성 변화를 찾지 못할 때나 50%를 기준으로 분류 시 검사자간 차이가 있어서 일 것으로 생각된다. 그 외 모든 소견들은 우수 이상의 일치도를 보였다.

그룹 1과 그룹 2 모두에서 대부분 우수 이상의 높은 일치도를 보였으나 그룹 1 보다는 그룹 2에서 우수 이상의 일치도를 보이는 초음파 소견이 더 많았다. 전공의와 전문의 사이의 갑상선 결절 초음파 소견에 대한 검사자간 일치도는 높고, 특히 교육을 더 많이 받은 전공의에서 조금 더 높음을 알 수 있었다.

그룹 1과 2 모두에서 전공의들의 갑상선 암에 대한 진단 정확도는 높았고, 전문의와 비교 시 통계학적으로 유의한 차이가 없었다. 일부 정확도는 Az 값이 1인 경우가 있었다. 검사를 시행한 결절의 수가 적은 것이 이유일 것으로 생각 된다.

이 연구에는 제한점들이 있다. 첫째, 이 연구는 후향적 연구로 연구 기간 동안 모든 갑상선 결절이 포함되지 않았다. 전문의가 세침흡입생검을 시행하고 전공의가 참관 후 기록을 한 결절만을 대상으로 하였다. 따라서 선택 바이어스가 있을 수 있고 결절의 수가 적다는 것이 제한점이다. 둘째, 초음파와 세침흡입생검을 시행한 갑상선 결절이 그룹내에서나 그룹간에서나 모두 다른 갑상선 결절이었다. 따라서 그룹 1과 2 전공의간 일치도를 직접 비교할 수 없었고, 갑상선 전문의와의 비교를 통해 간접적인 방법으로 비교를 하였다. 셋째, 양성, 악성의 판단 기준을 일부는 세침흡입생검으로 하였다. 세침흡입생검의 위 양성, 위음성이 존재 할 수 있다.

결론적으로, 전공의와 전문의간 갑상선 결절의 초음파 소견에 대한 일치도는 높았고 이는 교육을 통해 조금 더 향상 될 수 있다. 갑상선 암에 대한 초음파 소견의 진단 정확도 또한 높았다. 따라서 갑상선 결절의 초음파 소견은 초심자도 쉽게 배우고 익혀 이용할 수 있는 유용한 소견이다.

요 약

목적: 갑상선 초음파 교육을 받은 전공의를 대상으로 갑상선 전문의와 검사자간 일치도와 진단 정확도 변화를 알아 보고자 한다.

대상과 방법: 2007년 10월부터 2009년 6월까지 갑상선 초음파 파트에서 18명의 전공의 교육이 있었다. 한 번 교육을 받은 8명 전공의를 그룹 1, 두 번 교육을 받은 10명 전공의를 그룹 2로 분류 하였다. 559 갑상선 결절이 연구에 포함 되었다. 전문의가 초음파 시행 후, 초음파 소견에 대해 구성 성분, 변연부, 에코, 석회화, 모양, 최종 결론에 따라 전문의와 전공의가 따로 기록하였다. 사진을 보고 다시 검토 후, 전문의가 세침흡입생검을 시행하였다. 초음파 소견에 대해 전공의와 전문의간 일치도와 진단 정확도를 교육의 정도에 따라 비교하였다.

결과: 559 결절 중 187은 악성, 372는 양성으로 분류 되었다. 그룹 1의 전공의 1은 구성 성분에서 불량한 일치도를 보였다. 전공의 5, 6은 석회화에서, 전공의 4는 모양에서 중증도의 일치도를 보였다. 그룹 2의 전공의 1과 10의 구성 성분에서 중증도의 일치도를 보였다. 그룹 2에서 우수 이상의 일치도를 보이는 경우가 그룹 1 보다 많았다. 그룹 1, 2 모두에서 진단 정확도는 높았으며 전문의와 통계학적인 차이가 없었다.

결론: 전공의와 전문의간 초음파 소견의 검사자간 일치도와 갑상선 암에 대한 진단 정확도는 높았고, 이는 교육을 통해 향상 될 수 있다.

References

1. Hegeds L. Clinical practice. The thyroid nodule. *N Engl J Med* 2004;351:1764-1771
2. Tan GH, Gharib H. Thyroid incidentalomas: management approaches to nonpalpable nodules discovered incidentally on thyroid imaging. *Ann Intern Med* 1997;126:226-231
3. 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;23:1455-1464
4. Kim E, Park CS, Chung WY, et al. New sonographic criteria for recommending fine-needle aspiration biopsy of nonpalpable solid nodules of the thyroid. *AJR Am J Roentgenol* 2002;178:687-691
5. Moon HJ, Kwak JY, Kim MJ, Son EJ, Kim E. Can vascularity at power Doppler US help predict thyroid malignancy? *Radiology* 2010;255:260-269
6. Moon WJ, Jung SL, Lee JH, et al. Benign and malignant thyroid nodules: US differentiation-multicenter retrospective

- tive study. *Radiology* 2008;247:762-770
7. Papini E, Guglielmi R, Bianchini A, et al. Risk of malignancy in nonpalpable thyroid nodules: predictive value of ultrasound and color-Doppler features. *J Clin Endocrinol Metab* 2002;87:1941-1946
 8. Cooper DS, Doherty GM, Haugen BR, et al. Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. *Thyroid* 2009;19:1167-1214
 9. Choi SY, Kim EK, Kwak JY, Kim MJ, Son EJ. Inter/intraobserver variation in US assessment of thyroid nodules. *Thyroid* 2009;In Revision
 10. 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;22:1027-1031
 11. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-174
 12. Kim SH, Park CS, Jung SL, et al. Observer variability and the performance between faculties and residents: US criteria for benign and malignant thyroid nodules. *Korean J Radiol* 2010;11:149-155