

Otrzymano:  
20.01.2019  
Zaakceptowano:  
29.03.2019  
Opublikowano:  
30.09.2019

## Prospektywna analiza powtarzalności i odtwarzalności oceny obrazów ultrasonograficznych guzków tarczycy

### Prospective analysis of inter-observer and intra-observer variability in multi ultrasound descriptor assessment of thyroid nodules

Katarzyna Dobruch-Sobczak<sup>1,2</sup>, Bartosz Migda<sup>3</sup>, Agnieszka Krauze<sup>3</sup>, Krzysztof Mlosek<sup>3</sup>, Rafał Z. Słapa<sup>3</sup>, Paweł Wareluk<sup>3</sup>, Elwira Bakuła-Zalewska<sup>4</sup>, Zbigniew Adamczewski<sup>5,6</sup>, Andrzej Lewiński<sup>5,6</sup>, Wiesław Jakubowski<sup>3</sup>, Marek Dedecjus<sup>7</sup>

<sup>1</sup> Zakład Radiologii II, Centrum Onkologii – Instytut im. Marii Skłodowskiej-Curie w Warszawie, Warszawa, Polska

<sup>2</sup> Oddział Ultrasonografii, Instytut Podstawowych Problemów Techniki, PAN, Warszawa, Polska

<sup>3</sup> Zakład Diagnostyki Obrazowej, Warszawski Uniwersytet Medyczny, II Wydział Lekarski z Oddziałem Nauczania w Języku Angielskim oraz Oddziałem Fizjoterapii, Warszawa, Polska

<sup>4</sup> Zakład Patologii, Centrum Onkologii – Instytut im. Marii Skłodowskiej-Curie w Warszawie, Polska

<sup>5</sup> Katedra Endokrynologii i Chorób Metabolicznych, Uniwersytet Medyczny w Łodzi, Polska

<sup>6</sup> Instytut Centrum Zdrowia Matki Polki, Polska

<sup>7</sup> Zakład Medycyny Nuklearnej i Onkologii Endokrynologicznej, Centrum Onkologii – Instytut im. Marii Skłodowskiej-Curie w Warszawie, Warszawa, Polska

Adres do korespondencji: Bartosz Migda, Zakład Diagnostyki Obrazowej, Warszawski Uniwersytet Medyczny, II Wydział Lekarski z Oddziałem Nauczania w Języku Angielskim oraz Oddziałem Fizjoterapii, Warszawa, Polska; tel.: +48 22 326 58 10, fax: +48 22 326 54 79, e-mail: bartoszmigda@gmail.com

DOI: 10.15557/JoU.2019.0030

#### Słowa kluczowe

zmienność  
odtwarzalności,  
zmienność  
powtarzalności,  
guzek tarczycy,  
ultrasonografia,  
sonoelastografia

#### Keywords

inter-observer  
variability,  
intra-observer  
variability,  
thyroid nodule,  
ultrasound,  
sonoelastography

#### Abstract

**Aim:** The aim of this study was to evaluate the inter- and intra-observer variability and accuracy of ultrasound assessment of thyroid nodules using a descriptive lexicon. **Materials and methods:** A prospective study was performed on complete ultrasound examinations, including sonoelastography and color Doppler ultrasound of 18 patients with 20 thyroid nodules. A total of 20 records of thyroid nodules from these techniques were duplicated, numbered, and randomly arranged. Five radiologists assessed the recordings independently. Cohen Kappa and Fleiss Kappa statistics were used to determine the degree of intra- and inter-observer agreement. **Results:** Mean accuracy rates for all radiologists, for all ultrasound features, ranged from 82.7 to 87.8%. For B-mode and strain elastography, accuracies ranged from 65.0 to 100% and 47.4 to 86.8%, respectively. Concerning intra-observer variability, three radiologists demonstrated almost perfect agreement (the  $\kappa$ -value ranged from 0.81 to 0.86), and a substantial agreement was noted for the two remaining radiologists. The  $\kappa$ -values for inter-observer agreement ranged from 0.61 for macrocalcifications (substantial agreement) to 0.33 for Asteria four-point elastography scale criteria (fair agreement). **Conclusions:** The results suggest relatively good inter-observer and excellent intra-observer agreement in the assessment of thyroid nodules using ultrasound, and fair agreement in the case of strain elastography.

## Wstęp

Na całym świecie odnotowuje się stały wzrost częstości zachorowania na raka tarczycy<sup>(1)</sup>. Według American Thyroid Association (ATA) guzki tarczycy stanowią powszechny problem kliniczny i stwierdzane są u niemal 68% dorosłych poddanych badaniu. W 7–15% tych przypadków rozpoznaje się raka tarczycy<sup>(2)</sup>. W Polsce w 2015 roku zdiagnozowano 3529 nowych przypadków raka tarczycy. Roczny wskaźnik zapadalności wzrósł z 3,8 na 100 000 w 2000 roku do 9,2 na 100 000 w 2015 roku<sup>(3,4)</sup>.

Choć guzki tarczycy charakteryzują się powszechnym występowaniem, wykrycie tych zmian bez zastosowania metod obrazowania jest zazwyczaj trudne (jedynie 4–7% można wykryć w badaniu palpacyjnym)<sup>(5)</sup>. Zatem badanie ultrasonograficzne odgrywa ważną rolę w diagnostyce, jako nieinwazyjna, niedroga i szeroko dostępna technika, wykorzystywana do identyfikacji swoistych cech guzków i monitorowania biopsji aspiracyjnej cienkoigłowej celowanej (BACC)<sup>(5)</sup>.

Opublikowane badania, w tym wytyczne ATA, wykazały, że hipoechogeniczność, nieregularny obrys, obecność mikrozwapnień i kształt zmiany „wyższy niż szerszy” to charakterystyczne cechy obrazu w trybie B-mode (w skali szarości), o najwyższej swoistości w wykrywaniu złośliwych guzów tarczycy<sup>(6–8)</sup>. Jednak żadna z tych cech, jeśli rozpatrywana jest indywidualnie, nie dotyczy wyłącznie zmian złośliwych, a guzki łagodne z izolowaną nieprawidłowością są stosunkowo powszechne<sup>(2,9–11)</sup>.

W związku z tym opracowywane są nowe, nieinwazyjne metody obrazowania umożliwiające przeprowadzenie diagnostyki różnicowej zmian ogniskowych tarczycy. W ostatnim czasie coraz większą popularnością cieszy się sonoelastografia<sup>(12,13)</sup>.

Obecnie dostępne są dwa główne typy sonoelastografii: elastografia fali poprzecznej (*shear wave elastography*, SWE) oraz elastografia quasi-statyczna/odkształceń względnych (*strain elastography*, SE). Choć niektórzy autorzy sugerują przewagę SWE nad SE w stratyfikacji ryzyka guzków tarczycy z uwagi na niezależność od operatora, najnowsze metaanalizy niespodziewanie wskazują na większą dokładność diagnostyczną SE w porównaniu ze SWE<sup>(14,15)</sup>. Ponadto Dighe i wsp. opisali obecność artefaktów w badaniu metodą SWE oraz ich wpływ na końcowe wyniki badania<sup>(16)</sup>. Autorzy zasugerowali, że niemal 70% skanów SWE zawiera artefakty; ponad 18% uzyskiwanych skanów nie nadawało się do ostatecznej oceny, a ponad 43% artefaktów w obrazie SWE było zależnych od operatora<sup>(16)</sup>. Elastografia quasi-statyczna również ma swoje ograniczenia. Wyniki badania w znacznej mierze zależą od obecności zwapnień w guzkach (makrozwapnień i zwapnień brzeżnych), jak również ich lokalizacji (zmiany położone głęboko) oraz typu raka [rak brodawkowaty tarczycy (*papillary thyroid cancer*, PTC) jest zwykle bardziej podejrzany niż rak pęcherzykowy tarczycy (*follicular thyroid carcinoma*, FTC)]. Ocena ryzyka złośliwości guzka tarczycy z zastosowaniem elastografii quasi-statycznej

i obrazowania w skali szarości jest zależna od doświadczenia operatora i charakteryzuje się znaczną zmiennością wyników uzyskanych przez różnych badających<sup>(5,17–20)</sup>, ale metody te są pomocne w monitorowaniu pacjentów poddanych wcześniejszej BACC<sup>(21–23)</sup>.

Niewiele badań poświęcono ocenie powtarzalności i odtwarzalności obrazów USG, a jeszcze mniej prac dotyczy zgodności w przypadku sonoelastografii<sup>(5,17–20)</sup>. W związku z tym przeprowadziliśmy pełną analizę powtarzalności i odtwarzalności oceny w obrazowaniu B-mode i sonoelastografii.

## Materiał i metody

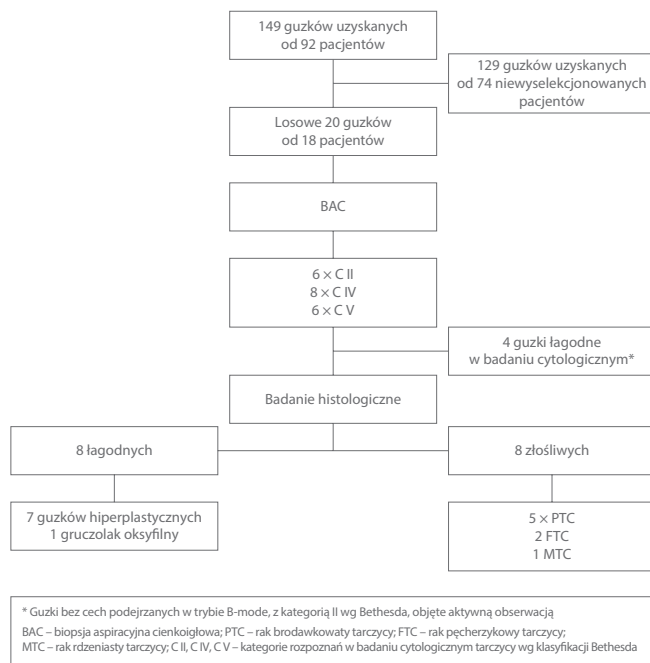
### Pacjenci

W badaniu przeprowadzono prospektywną ocenę obrazów B-mode i sonoelastografii. Wszyscy pacjenci udzieliли świadomej zgody na udział w badaniu. U wszystkich osób wykonano badanie USG tarczycy, a następnie BACC lub zabieg chirurgiczny. Protokół badania został zatwierdzony przez Komisję Rewizyjną Centrum Onkologii – Instytutu im. Marii Skłodowskiej-Curie w Warszawie (Polska). W okresie od lutego do listopada 2017 r. w Klinice Endokrynologii Onkologii i Medycyny Nuklearnej Centrum Onkologii – Instytutu im. Marii Skłodowskiej-Curie w Warszawie (Polska) włączono do badania i przebadano 149 guzków tarczycy u 92 pacjentów (22 mężczyzn i 70 kobiet). Z tej grupy wybrano losowo 18 pacjentów (4 mężczyzn, 14 kobiet) w wieku 21–78 lat, u których stwierdzono 20 guzków tarczycy, w tym 8 zmian złośliwych i 12 zmian o charakterze łagodnym. W przypadku wszystkich zmian złośliwych oraz 8 guzków o charakterze łagodnym potwierdzono rozpoznanie w pooperacyjnym badaniu histopatologicznym. Pozostałe 4 guzki sklasyfikowane w badaniu cytologicznym jako CII nie zostały zaklasyfikowane do zabiegu chirurgicznego, gdyż operowanie pacjentów przy braku wskazań byłoby nieetyczne. Osoby z tej grupy poddawano kontrolnemu badaniu USG w odstępach 6-miesięcznych (Ryc. 1).

Guzki włączane do badania stanowiły grupę zmian poddanych BACC pod kontrolą USG (zgodnie z wytycznymi polskich towarzystw narodowych<sup>(23)</sup>, opracowanych z inicjatywy Polskiej Grupy ds. Nowotworów Endokrynnych i ATA), a następnie operacji lub aktywnej obserwacji, w tym wielokrotnej BACC. Z ostatecznej analizy wykluczono zmiany czysto torbielowate, zwapnienia skorupkowate lub niediagnostyczne wyniki badania cytologicznego. Lekarze wykonujący badanie USG nie mieli wglądu do wyników badania cytologicznego i histopatologicznego.

### Badanie histopatologiczne

Tyroidektomię i BACC wykonano u 14 pacjentów, natomiast 4 pacjentów poddano samej BACC. Wyniki badań histopatologicznych i cytologicznych stanowiły punkty końcowe. W przypadku pacjentów z rozpoznaniem łagodnych



Ryc. 1. Schemat blokowy

guzków tarczycy w badaniu BACC kontrolne badanie USG wykonywano w odstępach 6 miesięcy. Biopsję cienkoigłową aspiracyjną celowaną przeprowadzano z zastosowaniem igieł 22–24 G, natomiast aspiraty utrwalono w 75-procentowym roztworze etanolu i barwiono hematoksyliną i eozyną (H&E). Wyniki biopsji zmian klasyfikowano według systemu Bethesda (I–IV)<sup>(24)</sup>. W przypadku stwierdzenia zmian kategorii CI (próbka niediagnostyczna, np. tylko płyn z torbieli, próbka bezkomórkowa), CIII (atypia o nieokreślonym znaczeniu/zmiana pęcherzykowa bliżej nieokreślona) oraz małych guzków CIV (<1 cm; podejrzenie nowotworu pęcherzykowego w przypadku guzków <1 cm) wykonywano powtórny biopsję cienkoigłową aspiracyjną celowaną. W miarę możliwości sugerowano określony histotyp. Wyniki badania cytologicznego (CV i CVI) były weryfikowane przez drugiego, niezależnego patologa. Próbkę chirurgiczną bezpośrednio utrwalano w 10-procentowej formalinie buforowanej. Reprezentatywne wycinki z pobranych próbek poddano obróbce i rutynowemu barwieniu H&E w celu przeprowadzenia badania histopatologicznego (mikroskopowego).

## Konwencjonalne badania w trybie B-mode i USG

Badania zostały przeprowadzone i ocenione przez pięciu radiologów (jednego z ośrodka onkologicznego i czterech z ośrodka klinicznego) z doświadczeniem w wykonywaniu badania USG w trybie B-mode i z opcją kolorowego dopplera (*color doppler ultrasonography*, CDUS) wynoszącym od 6 do 22 lat oraz z doświadczeniem w badaniu metodą sonoelastografii wynoszącym od 1 do 7 lat. Przed rozpoczęciem badania wszyscy radiolodzy zostali przeszkoleni w zakresie autorskiego leksykonu obejmującego następujące parametry: echostruktura; echogeniczność zmiany w porównaniu z mięszem tarczycy (Echo-Pa); dominująca

echogeniczność zmiany w porównaniu z mięszem tarczycy – w przypadku mieszanej echogeniczności dominujący objętościowo komponent (Echo-Pb); echogeniczność zmiany w porównaniu z mięśniami podgnykowymi (Echo-M); granice; objaw „halo”; naciek pozatarczycowy (badających poproszono o ustalenie, czy obecne jest modelowanie kształtu gruczołu tarczowego i torebki, czy obecne jest wykraczanie poza nią); makrozwapnienia; mikrozwapnienia; ocena elastyczności (według skali Asterii) – wszystkie parametry zestawiono w Tab. 1. Wszystkie badania przeprowadzono według tego samego, opisanego poniżej protokołu.

Badanie wykonywano w ułożeniu na wznak, uzyskując przekroje podłużne i poprzeczne z minimalnym uciskiem głowicą. Obrazowanie tarczycy prowadzono w kierunku góra–dół w przekroju poprzecznym oraz od granicy bocznej do przyśrodkowej (prztychawicz) w przekrojach podłużnych. Wymiary przednio-tylne, poprzeczne i podłużne płatów i guzków mierzono na pojedynczych obrazach, a następnie archiwizowano. Pozostałe cechy w trybie B-mode zawarte w leksykonie oraz parametry badania z opcją kolorowego dopplera i SE oceniano retrospektywnie na podstawie zarchiwizowanych pętli AVI i pojedynczych obrazów. Badanie z zastosowaniem kolorowego dopplera wykonywano we wszystkich przypadkach przy jednakowym ustawieniu skali (prędkość maksymalna 2,5 cm/s). Wzmocnienie w badaniu z opcją kolorowego dopplera dostosowywano indywidualnie do każdego pacjenta, uzyskując odpowiednią, najwyższą czułość bez obecności artefaktów „rozlewania się” (*blooming artifacts*). W przypadku badania metodą SE, z uwagi na fakt, że ucisk guzków powoduje zwiększenie ich sztywności, wszyscy radiolodzy unikali uciskania szyi głowicą w trakcie badania w celu zminimalizowania wyników fałszywie dodatnich. Badania USG w opcjach B-mode, kolorowego dopplera i SE wykonywano aparatem iU22 (Philips Medical Systems, Bothell, WA) wyposażonym w głowicę liniową o częstotliwości 5–12 MHz. Ocenę jakościową sonoelastografii oparto na kryteriach czteropunktowej skali według Asterii (Tab. 1)<sup>(25)</sup>. W badaniu USG i SE ocenie poddano cechy zmian przedstawione w Tab. 1. Z ostatecznej oceny wykluczono kształt zmiany (wyższy niż szerszy), gdyż ocena tego parametru jest bardziej obiektywna, z uwagi na fakt, że polega na porównaniu wymiarów guzka (wysokości i szerokości, mierzonych prospektywnie). Ze względu na charakter badania ocena powtarzalności i odtwarzalności w odniesieniu do tego parametru mogłaby zawyżyć końcowe wyniki.

## Ultrasonograficzna ocena guzków przez niezależnych badających

Spośród 149 zbadanych guzków tarczycy losowo wybrano 20. W tym celu wykorzystano odpowiednie formuły arkusza MS Excel. Uzyskane obrazy 20 guzków w opcji B-mode, kolorowego dopplera i SE zduplikowano – ostatecznie 40 skanów ponumerowano i ułożono losowo w końcowym zbiorze. Następnie pięciu radiologów wykonało ocenę tych samych zbiorów badań (pętli AVI oraz obrazów w pliku JPG), obejmujących poprzeczne i podłużne przekroje

**Tab. 1.** Wykaz ultrasonograficznych cech guzków tarczycy

Cechy USG	Skrót	Charakterystyka
Echostruktura	Struktura	Torbielowata Gąbczasta Mieszana torbielowata (≥50% obj. torbielowatej) Mieszana lita (≥50% obj. litej) Prawie całkowicie lita Lita Nie do określenia
Echogeniczność (w porównaniu z mięszem tarczycy)	Echo-Pa	Izoehogeniczny Mieszany Hipoehogeniczny
Dominująca echogeniczność (w porównaniu z mięszem tarczycy)	Echo-Pb	Hiperechogeniczny Izoehogeniczny Hipoehogeniczny
Echogeniczność (w porównaniu z mięśniami)	Echo-M	Hiperechogeniczny Izoehogeniczny Hipoehogeniczny
Granice	Granice	Dobrze odgraniczone Źle odgraniczone
Objaw „Halo”	„Halo”	Całkowity Częściowy Brak
Naciek pozatarczycowy	Torebka	Modelowanie tarczycy i torebki Inwazja poza torebkę tarczycy Brak
Makrozwapnienia (>1 mm)	Makro	Obecne Brak
Mikrozwapnienia (≤1 mm)	Mikro	Obecne Brak
Unaczynienie	Unaczynienie	Obwodowe Centralne Mieszane (centralne i obwodowe) Brak
Skala elastyczności wg Asterii	Skala Asterii	I – Elastyczność na całym badanym obszarze II – Elastyczność na znacznej części badanego obszaru III – Brak elastyczności na znacznej części badanego obszaru IV – Brak elastyczności na całym badanym obszarze
Pozaguzkowy miąższ tarczycy	Miąższ	Jednorodny Niejednorodny
Autoimmunologiczne zapalenie tarczycy	AT	Obecne Brak
Unaczynienie miąższu	Unaczynienie miąższu	Prawidłowe Skąpe Wzmoczone

płatów tarczycy w opcji B-mode oraz guzków tarczycy w opcji CDUS i SE.

### Analiza statystyczna

Wyniki dla wszystkich pięciu badających obliczono przy użyciu oprogramowania statystycznego Dell Statistica (wersja 13, software.dell.com, Dell Inc. 2016). Ogólną wartość  $\kappa$  oszacowano dla wszystkich obserwatorów. W celu określenia stopnia powtarzalności na podstawie zduplikowanych wyników tego samego pacjenta zastosowano współczynnik  $\kappa$  Cohena po skorygowaniu o zgodność przypadkową. Do oceny odtwarzalności zastosowano współczynnik  $\kappa$  Fleissa. W tym przypadku również dokonano korekcji pod kątem zgodności przypadkowej<sup>(26)</sup>.

Wartości  $\kappa$  interpretowano według Landisa i Kocha<sup>(27)</sup>, tj.  $\kappa < 0,00$  – słaba zgodność,  $\kappa = 0,00-0,20$  – niewielki stopień zgodności,  $\kappa = 0,21-0,40$  – dobra zgodność,  $\kappa = 0,41-0,60$  – umiarkowana zgodność,  $\kappa = 0,61-0,80$  – wysoka zgodność, a  $\kappa = 0,81-1,00$  – prawie doskonała zgodność.

Na koniec poddano ocenie i porównano dokładność wszystkich badających i określono modę dla każdego deskryptora w tym zbiorze danych. Wartość tę przyjęto za poprawną dla danego deskryptora. Badaczom, którzy zgodzili się z tą wartością, przyznano wartość dokładności 1; pozostali otrzymali wartość dokładności równą 0. Następnie obliczono wynik całkowitej dokładności niezależnie dla każdego obserwatora względem każdego deskryptora.

### Wyniki

Nasza losowo wybrana grupa składała się z 20 guzków stwierdzonych u 18 pacjentów (12 guzków łagodnych i 8 złośliwych). Maksymalna długość guzów wynosiła od 6 do 46 mm (średnia długość  $9,7 \pm 5,6$  mm). Pięć z tych zmian rozpoznano jako PTC, dwa jako FTC, a jedną zmianę jako raka rdzeniastego tarczycy (*medullary thyroid carcinoma*, MTC). W grupie pacjentów z guzkami łagodnymi 8 zmian zweryfikowano w badaniu histopatologicznym i większość (7/8) z nich stanowiły guzki rozrostowe (Ryc. 1).

Wyniki oceny dokładności pięciu radiologów przedstawiono w Tab. 2. Średnie wartości współczynnika dokładności dla wszystkich badających wynosiły 82,7–87,8%. Wszyscy radiolodzy uzyskali wartości współczynnika dokładności wynoszące 65,0–100% w przypadku obrazowania w opcji B-mode i 47,4–86,8% dla SE. Najwyższy poziom dokładności wśród badających odnotowano w odniesieniu do następujących cech: makrozwapnienia (90,0–100%), mikrozwapnienia (85,0–100%) oraz ocena wzoru echogeniczności w porównaniu z mięśniami podgnykowymi (87,5–95,0%). Dane dotyczące zmienności wyników powtarzalności i odtwarzalności w odniesieniu do cech uwidocznionych w obrazowaniu guzków tarczycy w trybie B-mode, CDUS i SE przedstawiono w Tab. 3.

Tab. 2. Ocena dokładności badającego

Opis	Dokładność (%)				
	Badający 1	Badający 2	Badający 3	Badający 4	Badający 5
Echostruktura	92,5	95,0	77,5	72,5	90,0
Echo-Pa	80,0	80,0	90,0	90,0	82,5
Echo-Pb	92,5	82,5	87,5	95,0	87,5
Echo-M	87,5	95,0	90,0	92,5	97,5
Granice	82,5	87,5	72,5	82,5	92,5
„Halo”	85,0	87,5	82,5	65,0	87,5
Torebka	87,5	85,0	80,0	92,5	92,5
Makro	100,0	92,5	97,5	90,0	97,5
Mikro	100,0	92,5	85,0	87,5	95,0
Unaczynienie	65,0	95,0	85,0	97,5	75,0
Skala Asterii	78,9	86,8	68,4	47,4	73,7
Miąższ	82,5	82,5	80,0	87,5	65,0
AT	77,5	80,0	92,5	97,5	100,0
Unaczynienie miąższu	75,0	87,5	70,0	77,5	85,0
Średnia	84,7	87,8	82,7	83,9	87,2

Trzej badający uzyskali niemal doskonałą zgodność powtarzalności: drugi, trzeci i czwarty badający uzyskali wartości  $\kappa$  wynoszące odpowiednio 0,82, 0,86 i 0,81. Wysoką zgodność w odniesieniu do wartości  $\kappa$  odnotowano również dla pierwszego i piątego badającego. Zgodność wyników odtwarzalności określona przez wartości  $\kappa$  wahała się od 0,61 w przypadku makrozwapnień (wysoka zgodność) do 0,33 według kryteriów Asterii (dobra zgodność).

Odtwarzalność w odniesieniu do większości cech USG cechowała się umiarkowanymi wartościami w przypadku oceny mikrozwapnień ( $\kappa = 0,57$ ) (Ryc. 2) oraz echostruktury ( $\kappa = 0,55$ ), echogeniczności – Echo-Pa, Echo-Pb

i Echo-M ( $\kappa = 0,48-0,50$ ), inwazji pozataarczycowej – Torebka ( $\kappa = 0,40$ ) (Ryc. 3A). W przypadku oceny unaczynienia odnotowano ogólną dobrą zgodność ( $\kappa = 0,34$ ). Średnia wartość zgodności między obserwatorami dla wszystkich cech ultrasonograficznych i sonoelastografii wynosiła 0,42, co odpowiada umiarkowanej zgodności (Ryc. 3 B).

## Omówienie

Ultrasonografia jest powszechnie przyjętą techniką obrazową, umożliwiającą wykrywanie guzków i zmian echostruktury tarczycy. W ciągu ostatniej dekady dokonano znacznych ulepszeń technologicznych głowic i aparatury ultrasonograficznej. Obrazowanie takich cech ultrasonograficznych jak sztywność zmian oraz dokładna ocena mikrozwapnień, unaczynienia czy granic zmiany pozwalają na lepszą ocenę ryzyka złośliwości i dały podstawę do opracowania kilku klasyfikacji TIRADS (Thyroid Imaging Reporting and Data System), choć żadna z nich nie znalazła zastosowania w praktyce klinicznej w Polsce<sup>(9,28-32)</sup>. Podstawowym celem naszego badania była ocena powtarzalności i odtwarzalności w odniesieniu do cech ultrasonograficznych i sonoelastograficznych, co stanowi pierwszy krok w kierunku zaproponowania klasyfikacji TIRADS.

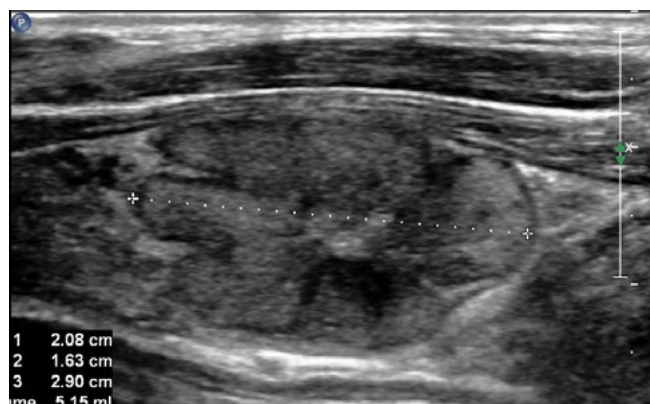
## Zwapnienia

W przedstawionym badaniu najwyższą zgodność uzyskano w przypadku oceny makrozwapnień: wartość  $\kappa$  wynosiła 0,61 dla odtwarzalności i 0,64-1,0 dla powtarzalności. W przypadku mikrozwapnień, które wykazują silniejszy związek ze złośliwym charakterem

Tab. 3. Ocena powtarzalności i odtwarzalności obserwatorów

Opis	Powtarzalność										Odtwarzalność
	Badający 1		Badający 2		Badający 3		Badający 4		Badający 5		
	Zgodność (%)	Wartość $\kappa$ (SE)	Zgodność (%)	Wartość $\kappa$ (SE)	Zgodność (%)	Wartość $\kappa$ (SE)	Zgodność (%)	Wartość $\kappa$ (SE)	Zgodność (%)	Wartość $\kappa$ (SE)	
Echostruktura	95,0	0,88 (0,12)	100,0	1,00 (0,00)	90,0	0,83 (0,11)	95,0	0,91 (0,09)	90,0	0,75 (0,17)	0,55 (0,04)
Echo-Pa	75,0	0,57 (0,16)	100,0	1,00 (0,00)	90,0	0,83 (0,11)	100,0	1,00 (0,00)	75,0	0,53 (0,17)	0,48 (0,04)
Echo-Pb	85,0	0,69 (0,16)	95,0	0,91 (0,09)	95,0	0,89 (0,11)	100,0	1,00 (0,00)	80,0	0,52 (0,21)	0,50 (0,05)
Echo-M	90,0	0,80 (0,14)	100,0	1,00 (0,00)	100,0	1,00 (0,00)	95,0	0,84 (0,16)	90,0	0,67 (0,18)	0,49 (0,04)
Granice	85,0	0,70 (0,16)	95,0	0,90 (0,10)	95,0	0,88 (0,12)	95,0	0,88 (0,12)	85,0	0,63 (0,20)	0,39 (0,05)
„Halo”	80,0	0,68 (0,13)	85,0	0,76 (0,12)	85,0	0,77 (0,12)	90,0	0,67 (0,19)	75,0	0,62 (0,14)	0,41 (0,04)
Torebka	80,0	0,64 (0,17)	95,0	0,91 (0,09)	85,0	0,72 (0,15)	80,0	0,59 (0,18)	75,0	0,50 (0,19)	0,40 (0,04)
Makro	100,0	1,00 (0,00)	95,0	0,83 (0,17)	95,0	0,64 (0,33)	100,0	1,00 (0,00)	95,0	0,64 (0,33)	0,61 (0,05)
Mikro	95,0	0,89 (0,11)	95,0	0,90 (0,10)	90,0	0,77 (0,15)	90,0	0,78 (0,14)	90,0	0,74 (0,17)	0,57 (0,05)
Unaczynienie	90,0	0,86 (0,10)	90,0	0,74 (0,16)	100,0	1,00 (0,00)	95,0	0,85 (0,13)	95,0	0,87 (0,13)	0,34 (0,03)
Skala Asterii	78,9	0,71 (0,13)	78,9	0,70 (0,13)	94,7	0,92 (0,08)	73,7	0,61 (0,15)	73,7	0,65 (0,13)	0,33 (0,03)
Miąższ	75,0	0,50 (0,19)	85,0	0,69 (0,16)	100,0	1,00 (0,00)	95,0	0,89 (0,10)	90,0	0,69 (0,20)	0,40 (0,05)
AT	95,0	0,88 (0,12)	80,0	0,47 (0,23)	95,0	*	95,0	0,64 (0,33)	100,0	1,00 (0,00)	0,25 (0,05)
Unaczynienie miąższu	80,0	0,64 (0,16)	85,0	0,66 (0,17)	95,0	0,92 (0,08)	85,0	0,71 (0,15)	75,0	0,17 (0,26)	0,18 (0,4)
Średnia	86,0	0,74	91,4	0,82	93,6	0,86	92,0	0,81	84,9	0,64	0,42

\* Struktura danych nie pozwoliła na obliczenie wartości  $\kappa$  i SE.



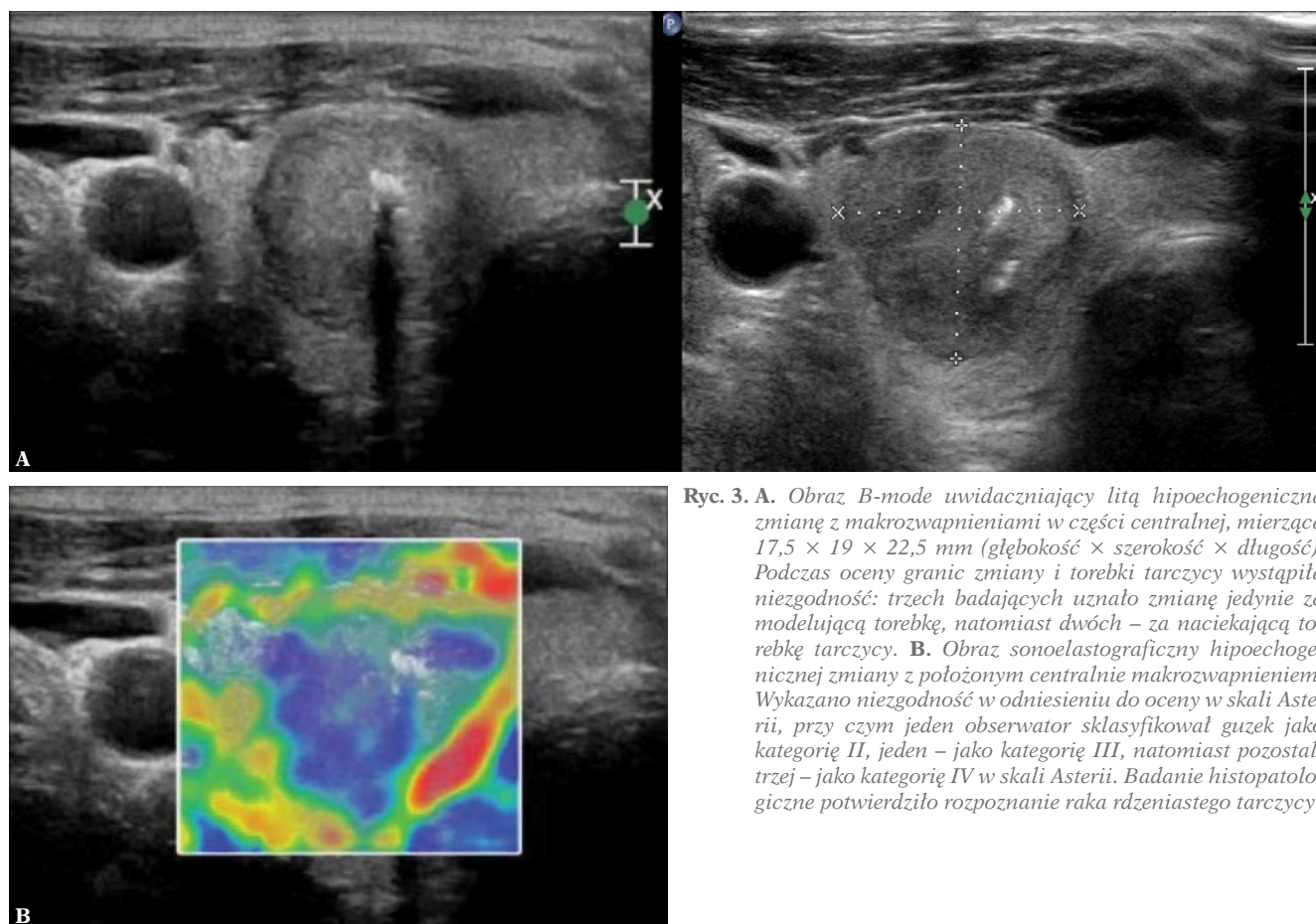
**Ryc. 2.** Obraz B-mode uwidaczniający litą, niejednorodną hipoechogeniczną zmianę o zrazikowych granicach, mierzącą 21 × 17 × 28,5 mm (głębokość × szerokość × długość). Podczas oceny mikrozwapnień wystąpiła niezgodność między badającymi: trzech uzyskało wynik ujemny, natomiast dwóch – wynik dodatni. Rozpoznanie histopatologiczne: rak pęcherzykowy tarczycy

zmian niż makrozwapnienia, uzyskano umiarkowaną zgodność<sup>(33)</sup>. Z tego względu były one oceniane osobno. Nasze wyniki są zbliżone z wynikami, które uzyskali Park i wsp.<sup>(18)</sup>, stosując takie same kryteria: mikrozwapnienia ≤1 mm i makrozwapnienia >1 mm.

W tym badaniu pięciu radiologów z doświadczeniem w ocenie guzków tarczycy wynoszącym od 1 roku do 6 lat uzyskali wartości  $\kappa = 0,40$  w przypadku makrozwapnień i 0,54 w przypadku mikrozwapnień.

W innym badaniu, w którym Park i wsp.<sup>(20)</sup> oceniali jedynie przypadki raka tarczycy (51 z 52 zmian rozpoznano jako raka brodawkowatego tarczycy), obecność zwapnień odnotowano w ponad połowie guzków (w zależności od badającego od 34 do 42 spośród 52 guzków; 65,4–80,7%). Autorzy ci uzyskali podobne wyniki do naszych, przy czym wartości  $\kappa$  wynosiły od 0,47 do 0,62 dla wszystkich typów zwapnień. W przedstawionym badaniu obecność mikrozwapnień stwierdzono w 7 spośród 8 nowotworów tarczycy, natomiast makrozwapnienia były obecne w 2 spośród 8 zmian nowotworowych. Ponadto współczynnik zgodności w odniesieniu do oceny zwapnień w przedstawionym badaniu był porównywalny z wynikami Choi i wsp.<sup>(5)</sup> Obecność mikrozwapnień w obrazie ultrasonograficznym wskazuje na potrzebę przeprowadzenia biopsji, a przede wszystkim dokładnej oceny cytologicznej<sup>(6)</sup>.

W obecnym badaniu retrospektywna analiza uzyskanych wyników wykazała brak zgodności w zakresie oceny mikrozwapnień w przypadku guzków, które były w przewadze normoechogeniczne lub zawierały komponenty hiper-echogeniczne (w przypadku echogeniczności mieszanej,



**Ryc. 3. A.** Obraz B-mode uwidaczniający litą hipoechogeniczną zmianę z makrozwapnieniami w części centralnej, mierzącą 17,5 × 19 × 22,5 mm (głębokość × szerokość × długość). Podczas oceny granic zmiany i torebki tarczycy wystąpiła niezgodność: trzech badających uznało zmianę jedynie za modelującą torebkę, natomiast dwóch – za naciekającą torebkę tarczycy. **B.** Obraz sonoelastograficzny hipoechogenicznej zmiany z położonym centralnie makrozwapnieniem. Wykazano niezgodność w odniesieniu do oceny w skali Asterii, przy czym jeden obserwator sklasyfikował guzek jako kategorię II, jeden – jako kategorię III, natomiast pozostali trzej – jako kategorię IV w skali Asterii. Badanie histopatologiczne potwierdziło rozpoznanie raka rdzeniastego tarczycy

w której mikrozwapnienia były obecne w komponencie hiperechogenicznej). Mogło to wpłynąć na kontrast między punktowym odbiciem  $<1$  mm a otaczającymi częściami litymi guzka. Niestety taki brak zgodności odnotowano w przypadku trzech zmian złośliwych: raka brodawkowatego tarczycy, wariantu pęcherzykowego raka brodawkowatego tarczycy oraz raka pęcherzykowego tarczycy (Ryc. 2). Wariant pęcherzykowy raka brodawkowatego tarczycy oraz rak pęcherzykowy tarczycy były normoechogeniczne, co mogło zmniejszyć wyżej wspomniany kontrast. Rak brodawkowaty tarczycy był hipoechogeniczny, jednak jego wymiar wynosił  $<10$  mm, co mogło stanowić kolejne ograniczenie w ocenie mikrozwapnień.

## Echogeniczność

W celu oceny echogeniczności guzka porównaliśmy go z mięszem tarczycy lub mięśniami podgnykowymi bądź wykorzystaliśmy dominującą echogeniczność. Analiza odtwarzalności wykazała umiarkowaną zgodność (wartości  $\kappa$  wynosiły od 0,48 do 0,5). Taki wynik można po części wyjaśnić złożoną echogenicznością guzów tarczycy oraz stanowiącego tło mięszu. Z analizy danych wynika, że poza złożoną echogenicznością ważna jest także struktura guzka. Brak zgodności częściej dotyczył guzków o mieszanej, lito-płynowej echostrukturze. Wielkość zmian również nie pozostawała bez znaczenia. Brak zgodności w odniesieniu do echogeniczności w porównaniu z mięszem częściej odnotowywano w przypadku dużych zmian wypełniających cały płat w stosunku do mniejszych guzków. Mogło to wynikać z mniejszej ilości otaczającego mięszu do porównania. Choi i wsp. odnotowali dobrą zgodność wyników oceny tych cech (wartości  $\kappa$  wynosiły 0,34 dla pierwszego i 0,45 dla drugiego badającego), które podzielili na cztery podkategorie: hiperechogeniczne, izoechogeniczne, hipoechogeniczne oraz wyraźnie hipoechogeniczne. Z drugiej strony, w naszym badaniu odnotowano niemal idealną zgodność powtarzalności (wartości  $\kappa$  od 0,83 do 1), pomimo bardzo szczegółowej definicji tej cechy, której przypisano siedem możliwych charakterystyk (Tab. 1).

## Granice zmiany

Dokładna ocena granic zmiany stanowi ważny parametr przy szacowaniu ryzyka złośliwości. W diagnostyce różnicowej guzków łagodnych i złośliwych zmiany o odgraniczonych brzegach mają częściej charakter łagodny. Jednak cecha ta wykazuje niską czułość, gdyż odgraniczone brzegi obserwuje się w 33 do 93% nowotworów złośliwych<sup>(34)</sup>. Heterogeniczność zmiany lub nakładanie się na siebie granic guzków utrudniają ocenę ich obrysu. Wyniki przedstawione przez inne grupy badaczy wskazują na wysoki stopień zmienności odtwarzalności w odniesieniu do oceny granic zmiany<sup>(11)</sup>. W naszym badaniu granice zmiany opisywano jako dobrze odgraniczone lub źle odgraniczone (zrazikowe, spikularne, kanciaste, poszarpane). Ocena tego parametru wiązała się z najniższym poziomem zgodności odtwarzalności ( $\kappa = 0,39$ ) i zadowalającym poziomem zgodności powtarzalności ( $\kappa = 0,65$ – $0,90$ ). Choi i wsp.

oraz Park i wsp. stosowali takie same kategorie dla obrysu zmian i otrzymali podobne wyniki, przy czym wartości  $\kappa$  wynosiły odpowiednio 0,42 i 0,03–0,29<sup>(5,20)</sup>. W naszym badaniu brak zgodności dotyczył w większym stopniu guzków przylegających do torebki tarczycy oraz guzków położonych między cieśnią a płatem lub modelujących torebkę. Różnice te mogą wynikać z nierównej grubości torebki tarczycy pozostającej w kontakcie z guzkiem.

## Objaw „halo” i naciek pozatarczycowy

Kolejnymi cechami poddanymi ocenie były objaw „halo” i naciek pozatarczycowy. W tym przypadku odnotowano umiarkowany poziom zgodności odtwarzalności, świadczący o ograniczonej dokładności oceny tych cech. Park i wsp. przedstawili jeszcze mniej korzystne wyniki, uzyskując jedynie dobry poziom zgodności ( $\kappa = 0,32$ ) w ocenie nacieku na torebkę<sup>(18)</sup>. W prezentowanym badaniu w odniesieniu do objawu „halo” i nacieku na torebkę brak zgodności wykazano przede wszystkim w przypadku dużych zmian, które uwypuklały torebkę tarczycy, pozostawały w kontakcie z torebką (Ryc. 3 A) lub stanowiły część konglomeratu guzków.

## Elastografia odkształceń względnych

W naszym badaniu określenie sztywności zmian z zastosowaniem 4-stopniowej skali było trudnym zadaniem dla wszystkich badających, na co wskazuje jedynie dobry poziom zgodności. Czterech radiologów z kilkuletnim doświadczeniem w wykonywaniu badania SE osiągnęło poziom dokładności rzędu 68,4 do 86,8%. Jeden radiolog z zaledwie rocznym doświadczeniem osiągnął poziom dokładności 47,4% (Ryc. 3 B). Odnotowano dobry poziom zgodności odtwarzalności ( $\kappa = 0,33$ ). Mogło to wynikać z różnic w poziomie doświadczenia. Dane z piśmiennictwa wskazują na różnice między ośrodkami badawczymi<sup>(19,20)</sup>. Friedrich-Rust i wsp. zastosowali tę samą 4-stopniową skalę w ilościowej ocenie SE i uzyskali wysoki poziom zgodności między trzema badającymi ( $\kappa = 0,66$ ). Natomiast Park i wsp. uzyskali tylko niewielką zgodność między badającymi dla tej samej techniki ( $\kappa = 0,08$ – $0,22$ )<sup>(20)</sup>. Jednak przeprowadzone w 2010 roku metaanalizy dotyczące badań SE wykazały wysokie uogólnione wartości czułości i specyficzności w przypadku odpowiednio 92% i 90% względem ostatecznego rozpoznania<sup>(34)</sup>, co potwierdziło skuteczność tej metody, która mogłaby w naszej opinii stanowić ważne narzędzie pomocnicze dla doświadczonego radiologa. Wszyscy uczestnicy naszego badania oceniali te same kopie oryginalnych plików (pętle video AVI) wraz z obrazami w trybie B-mode. Stąd wyższy wynik oceny dokładności w porównaniu z nieruchomymi obrazami, które oceniali w swoim badaniu Park i wsp. Tylko dobra zgodność w przypadku badania SE może wynikać z mniejszego doświadczenia jednego z badających w stosowaniu tej techniki. Wyniki naszej analizy sugerują potrzebę dyskusji na temat tego, czy SE, która nadal stanowi nową i rzadko stosowaną technikę, powinna zostać uwzględniona w ostatecznym leksykonie.

## Ograniczenia badania

Prezentowane badanie ma pewne ograniczenia. Włączono do niego pacjentów z Oddziału Medycyny Nuklearnej i Onkologii Endokrynologicznej ze wstępnym rozpoznaniem podejrzanych guzków lub z rozpoznaniem raka. Zatem grupa pacjentów różniła się od ogólnej populacji objętej badaniem przesiewowym; udział zmian złośliwych w tej grupie wynosił 45%. Stanowi to ograniczenie większości badań prowadzonych w ośrodkach endokrynologicznych i onkologicznych, gdzie odsetek przypadków nowotworów złośliwych jest wysoki. Zaproponowany leksykon był bardzo szczegółowy i pomimo uprzedniego przeszkolenia wszystkich radiologów doszło do pewnych nieporozumień. Wyniki prezentowanego badania wskazują na zbyt dużą liczbę cech ultrasonograficznych do oceny guzków, które powinny zostać poddane ponownej krytycznej ewaluacji przed włączeniem do kolejnych badań. W naszym badaniu stosowano elastografię odkształceń względnych, zależną od operatora i obarczoną pewnymi ograniczeniami: ułożenie sondy w stosunku do tętnicy szyjnej wspólnej (*common carotid artery*, CCA) – więcej artefaktów przy ułożeniu sondy w przekroju poprzecznym blisko CCA; ucisk sondą; obecność zwapnień brzeżnych lub licznych makrozwapnień pokrywających guzek; elementy płynowe w guzku. Jednak w przypadku SWE, którą uznaje się za bardziej niezależną, najnowsze doniesienia wskazują, że technika ta również jest obciążona zależnymi od operatora artefaktami i ograniczeniami<sup>(16)</sup>. Ponadto w tym badaniu stosowano jeden aparat USG i nie porównano wyników badania SE przeprowadzonych z użyciem różnych systemów. Można przyjąć założenie, że stosowanie SE z użyciem różnych platform mogłoby wpłynąć na różnice w wynikach końcowych, jednak należałoby to zweryfikować w badaniach prospektywnych. Zatem podczas opracowywania leksykonu TIRADS trzeba by uwzględnić oprogramowanie oraz rodzaj aparatury USG.

## Piśmiennictwo

- Guth S, Theune U, Aberle J, Galach A, Bamberger CM: Very high prevalence of thyroid nodules detected by high frequency (13 MHz) ultrasound examination. *Eur J Clin Invest* 2009; 39: 699–706.
- Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE *et al.*: 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association guidelines task force on thyroid nodules and differentiated thyroid cancer. *Thyroid* 2016; 26: 1–133.
- Didkowska J, Wojciechowska U, Olasek P: Nowotwory złośliwe w Polsce w 2015 roku. Centrum Onkologii – Instytut im. Marii Skłodowskiej-Curie, Warszawa 2017.
- Instytut ZEiPNCO: Krajowy Rejestr Nowotworów. Available from: [http://onkologia.org.pl/raporty/#tabela\\_nowotwor](http://onkologia.org.pl/raporty/#tabela_nowotwor).
- Choi SH, Kim EK, Kwak JY, Kim MJ, Son EJ: Interobserver and intraobserver variations in ultrasound assessment of thyroid nodules. *Thyroid* 2010; 20: 167–172.
- Brito JP, Gionfriddo MR, Nofal A, Boehmer KR, Leppin AL, Reading C *et al.*: The accuracy of thyroid nodule ultrasound to predict thyroid cancer: systematic review and meta-analysis. *J Clin Endocrinol Metab* 2013; 99: 1253–1263.
- Campanella P, Ianni F, Rota CA, Corsello SM, Pontecorvi A: Quantification of cancer risk of each clinical and ultrasonographic suspicious

## Podsumowanie

W prezentowanym badaniu pięciu radiologów (każdy z ponad 6-letnim doświadczeniem w obrazowaniu w opcji B-mode) dokonało oceny 40 guzków tarczycy. Wykazano stosunkowo dobrą odtwarzalność oraz doskonałą powtarzalność w odniesieniu do oceny guzków tarczycy w opcji B-mode oraz dobrą powtarzalność w przypadku sonoelastografii. Najwyższy poziom niezgodności dotyczył oceny nacieku pozatarczycowego, objawu „halo” oraz granic dużych zmian, zwłaszcza tych wypełniających większość płata tarczycy i/lub znajdujących się podtorebkowo. W przypadku makrozwapnień różnice najczęściej dotyczyły guzków normoechogenicznych lub guzków z komponentą hiperechogeniczną.

## Wnioski

Radiolodzy powinni zachować czujność podczas oceny granic zmian oraz nacieku na torebkę tarczycy w przypadku dużych guzków wypełniających znaczną część płata lub umiejscowionych w pobliżu torebki gruczołu, jak również podczas oceny mikrozwapnień w guzkach normoechogenicznych lub z komponentą hiperechogeniczną.

Łącznie wyniki wskazują na dobry poziom odtwarzalności i doskonały poziom powtarzalności w przypadku badania USG guzków tarczycy, a także dobry poziom powtarzalności w przypadku sonoelastografii.

## Konflikt interesów

*Autorzy nie zgłaszają żadnych finansowych ani osobistych powiązań z innymi osobami lub organizacjami, które mogłyby negatywnie wpłynąć na treść publikacji oraz rościć sobie do niej prawo.*

- feature of thyroid nodules: A systematic review and meta-analysis. *Eur J Endocrinol* 2014; 170: R203–R211.
- Remonti LR, Kramer CK, Leitao CB, Pinto LCF, Gross JL: Thyroid ultrasound features and risk of carcinoma: A systematic review and meta-analysis of observational studies. *Thyroid* 2015; 25: 538–550.
- Kwak JY, Han KH, Yoon JH, Moon HJ, Son EJ, Park SH *et al.*: Thyroid imaging reporting and data system for US features of nodules: A step in establishing better stratification of cancer risk. *Radiology* 2011; 260: 892–899.
- Salmashloğlu A, Erbil Y, Dural C, İşsever H, Kapran Y, Özarmağan S *et al.*: Predictive value of sonographic features in preoperative evaluation of malignant thyroid nodules in a multinodular goiter. *World J Surg* 2008; 32: 1948–1954.
- Moon WJ, Jung SL, Lee JH, Na DG, Baek JH, Lee YH *et al.*: Benign and malignant thyroid nodules: US differentiation – multicenter retrospective study. *Radiology* 2008; 247: 762–770.
- Szczepanek-Parulska E, Woliński K, Stangierski A, Gurgul E, Biczysko M, Majewski P *et al.*: Comparison of diagnostic value of conventional ultrasonography and shear wave elastography in the prediction of thyroid lesions malignancy. *PLoS One* 2013; 8: e81532.
- Szczepanek-Parulska E, Woliński K, Stangierski A, Gurgul E, Ruchała M: Biochemical and ultrasonographic parameters influencing thyroid nodules elasticity. *Endocrine* 2014; 47: 519–527.



14. Tian W, Hao S, Gao B, Jiang Y, Zhang X, Zhang S *et al.*: Comparing the diagnostic accuracy of RTE and SWE in differentiating malignant thyroid nodules from benign ones: a meta-analysis. *Cell Biochem* 2016; 39: 2451–2463.
15. Hu X, Liu Y, Qian L: Diagnostic potential of real-time elastography (RTE) and shear wave elastography (SWE) to differentiate benign and malignant thyroid nodules: A systematic review and meta-analysis. *Medicine* 2017; 96: e8282.
16. Dighe M, Hippe DS, Thiel J: Artifacts in shear wave elastography images of thyroid nodules. *Ultrasound Med Biol* 2018; 44: 1170–1176.
17. Moon HJ, Yoon JH, Kwak JY, Chung WY, Nam KH, Jeong JJ *et al.*: Positive predictive value and interobserver variability of preoperative staging sonography for thyroid carcinoma. *AJR Am J Roentgenol* 2011; 197: W324–W330.
18. Park CS, Kim SH, Jung SL, Kang BJ, Kim JY, Choi JJ *et al.*: Observer variability in the sonographic evaluation of thyroid nodules. *J Clin Ultrasound* 2010; 38: 287–293.
19. Friedrich-Rust M, Meyer G, Dauth N, Berner C, Bogdanou D, Herrmann E *et al.*: Interobserver agreement of Thyroid Imaging Reporting and Data System (TIRADS) and strain elastography for the assessment of thyroid nodules. *PLoS One* 2013; 8: e77927.
20. Park SH, Kim SJ, Kim EK, Kim MJ, Son EJ, Kwak JY: Interobserver agreement in assessing the sonographic and elastographic features of malignant thyroid nodules. *AJR Am J Roentgenol* 2009; 193: W416–W423.
21. Cosgrove D, Piscaglia F, Bamber J, Bojunga J, Correas JM, Gilja O *et al.*: EFSUMB guidelines and recommendations on the clinical use of ultrasound elastography. Part 2: clinical applications. *Ultraschall Med* 2013; 34: 238–253.
22. Shiina T, Nightingale KR, Palmeri ML, Hall TJ, Bamber JC, Barr RG *et al.*: WFUMB guidelines and recommendations for clinical use of ultrasound elastography: Part 1: basic principles and terminology. *Ultrasound Med Biol* 2015; 41: 1126–1147.
23. Jarzab B, Dedecjus M, Słowińska-Klencka D, Lewiński A, Adamczewski Z, Aniński R *et al.*: Guidelines of Polish National Societies Diagnostics and Treatment of Thyroid Carcinoma. 2018 Update. *Endokrynol Pol* 2018; 69: 34–74.
24. Straccia P, Rossi ED, Bizzarro T, Brunelli C, Cianfrini F, Damiani D *et al.*: A meta-analytic review of the Bethesda System for Reporting Thyroid Cytopathology: has the rate of malignancy in indeterminate lesions been underestimated? *Cancer Cytopathol* 2015; 123: 713–722.
25. Asteria C, Giovanardi A, Pizzocaro A, Cozzaglio L, Morabito A, Somalvico F *et al.*: US-elastography in the differential diagnosis of benign and malignant thyroid nodules. *Thyroid* 2008; 18: 523–531.
26. Fleiss JL, Levin B, Paik MC: *Statistical Methods for Rates and Proportions*. John Wiley & Sons 2013.
27. Landis JR, Koch GG: The measurement of observer agreement for categorical data. *Biometrics* 1977; 30: 159–174.
28. Horvath E, Majlis S, Rossi R, Franco C, Niedmann JP, Castro A *et al.*: An ultrasonogram reporting system for thyroid nodules stratifying cancer risk for clinical management. *J Clin Endocrinol Metab* 2009; 94: 1748–1751.
29. Russ G, Royer B, Bigorgne C, Rouxel A, Bienvenu-Perrard M, Leenhardt L: Prospective evaluation of thyroid imaging reporting and data system on 4550 nodules with and without elastography. *Eur J Endocrinol* 2013; 168: 649–655.
30. Russ G, Bonnema SJ, Erdogan MF, Durante C, Ngu R, Leenhardt L: European Thyroid Association guidelines for ultrasound malignancy risk stratification of thyroid nodules in adults: the EU-TIRADS. *Eur Thyroid J* 2017; 6: 225–237.
31. Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teefey SA *et al.*: ACR thyroid imaging, reporting and data system (TI-RADS): white paper of the ACR TI-RADS committee. *J Am Coll Radiol* 2017; 14: 587–595.
32. Migda B, Migda M, Migda MS, Słapa RZ: Use of the Kwak Thyroid Image Reporting and Data System (K-TIRADS) in differential diagnosis of thyroid nodules: systematic review and meta-analysis. *Eur Radiol* 2018; 28: 2380–2388.
33. Moon HJ, Sung JM, Kim EK, Yoon JH, Youk JH, Kwak JY: Diagnostic performance of gray-scale US and elastography in solid thyroid nodules. *Radiology* 2012; 262: 1002–1013.
34. Bojunga J, Herrmann E, Meyer G, Weber S, Zeuzem S, Friedrich-Rust M: Real-time elastography for the differentiation of benign and malignant thyroid nodules: a meta-analysis. *Thyroid* 2010; 20: 1145–1150.