











Otrzymano:
30.11.2021
Zaakceptowano:
27.01.2022
Opublikowano:
13.04.2022

Podsumowanie metaanaliz badań uwzględniających klasyfikacje TIRADS (EU-TIRADS, ACR-TIRADS oraz K-TIRADS) w szacowaniu potencjału złośliwości zmian ogniskowych gruczołu tarczowego

Summary of meta-analyses of studies involving TIRADS classifications (EU-TIRADS, ACR-TIRADS, and K-TIRADS) in evaluating the malignant potential of focal lesions of the thyroid gland

Katarzyna Dobruch-Sobczak^{*1} , Zbigniew Adamczewski^{*2} ,
Marek Dedejusz³ , Andrzej Lewiński^{4,5} , Bartosz Migda^{*6} ,
Marek Ruchała⁷ , Anna Skowrońska-Szcześniak^{*8} ,
Ewelina Szczepanek-Parulska^{*7} , Klaudia Zajkowska^{*3} , Agnieszka Żyłka^{*3} 

¹ Zakład Radiologii II, Narodowy Instytut Onkologii im. Marii Skłodowskiej-Curie, Państwowy Instytut Badawczy, Warszawa, Polska

² Zakład Medycyny Nuklearnej, Uniwersytet Medyczny w Łodzi, Łódź, Polska

³ Klinika Endokrynologii Onkologicznej i Medycyny Nuklearnej, Narodowy Instytut Onkologii im. Marii Skłodowskiej-Curie, Państwowy Instytut Badawczy, Warszawa, Polska

⁴ Klinika Endokrynologii i Chorób Metabolicznych, Instytut Centrum Zdrowia Matki Polki, Łódź, Polska

⁵ Klinika Endokrynologii i Chorób Metabolicznych, Uniwersytet Medyczny w Łodzi, Łódź, Polska

⁶ Pracownia Diagnostyki Ultrasonograficznej, Zakład Radiologii Pediatrycznej, Wydział Medyczny, Warszawski Uniwersytet Medyczny, Warszawa, Polska

⁷ Katedra i Klinika Endokrynologii, Przemiany Materii i Chorób Wewnętrznych, Uniwersytet Medyczny w Poznaniu, Poznań, Polska

⁸ Klinika Chorób Wewnętrznych i Endokrynologii, Warszawski Uniwersytet Medyczny, Warszawa, Polska

* Autorzy w równym stopniu przyczynili się do powstania tej pracy

Autor do korespondencji: Katarzyna Dobruch-Sobczak; e-mail: katarzyna.dobruch-sobczak@pib-nio.pl

DOI: 10.15557/JoU.2022.0020

Słowa kluczowe

tarczycy;
klasyfikacje
ultrasonograficzne;
EU-TIRADS;
ACR-TIRADS;
K-TIRADS; BACC

Keywords

thyroid;
ultrasound
classifications;
EU-TIRADS;
ACR-TIRADS;
K-TIRADS; FNAB

Abstract

Numerous scientific societies around the world have published their TIRADS (Thyroid Imaging Reporting and Data System) classifications that evaluate the risk of malignancy of focal thyroid lesions, presenting different ultrasound features for each category and lesion size thresholds to determine eligibility for biopsy. The use of such risk estimation systems in focal thyroid lesions facilitates the reporting of thyroid ultrasound findings and improves the qualification of focal lesions for fine-needle aspiration biopsy (FNAB). In this publication, the three most popular TIRADS classifications, European – EU-TIRADS, Korean – K-TIRADS, and developed by the American Society of Radiology – ACR-TIRADS, are presented and discussed based on a literature review. The results of available head-to-head statistical analyses comparing the classifications are also presented. The advantage of the EU-TIRADS and K-TIRADS systems is that they include only the most important ultrasound features, so their application is not time-consuming, and the scores are easy to incorporate into clinical practice. ACR-TIRADS, unlike other scales, is based on a unique classification system and represents the most comprehensive classification. Each of the

■ five categories of ultrasound features – morphology, echogenicity, shape, margins, microcalcifications – are evaluated and assigned a score from 0 to 3, with a higher score being associated with a higher risk of cancer. Based on the available data, the greatest benefit has been demonstrated for the ACR-TIRADS classification, which also has implications for minimising the number of unnecessary FNABs. However, limitations related to the heterogeneity of the groups analysed in the study, including differences in the populations studied, inclusion criteria, proportions of patients of either sexes, and the number of malignant lesions analysed, should also be taken into account.

Wstęp

Badanie ultrasonograficzne (USG) to podstawowe narzędzie diagnostyczne w ocenie zmian ogniskowych tarczycy; jest badaniem nieinwazyjnym oraz szeroko dostępnym. Podstawowym ograniczeniem tej metody jest subiektywność w ocenie cech ultrasonograficznych oraz brak możliwości wizualizacji struktur położonych zamostkowo⁽¹⁾. Cechuje je stosunkowo niska czułość diagnostyczna dla poszczególnych cech badania USG w szacowaniu złośliwości zmian ogniskowych (27–63%) przy wysokiej swoistości opisywanej dla cech, takich jak: mikrozwapnienia (87,8%), centralne unaczynienie (78%), nieregularne brzegi (83,2%), kształt wyższy niż szerszy (96%), zastosowanie elastografii (86,2%)⁽²⁾. W 2009 roku Horvath i wsp., wzorując się na systemie klasyfikacji obrazów gruczołu piersiowego BIRADS (Breast Imaging Reporting and Data System), zaproponowali jako pierwsi podobną klasyfikację dotyczącą obrazowania w tarczycy – TIRADS (Thyroid Imaging Reporting and Data System). Miała ona na celu wprowadzenie leksykonu cech sonograficznych zmian ogniskowych tarczycy, kategoryzację wyników oraz określenie dalszego postępowania diagnostyczno-terapeutycznego, w tym wskazań do biopsji aspiracyjnej cienkoigłowej (BACC)⁽³⁾. Biopsja aspiracyjna cienkoigłowa celowana stanowi istotne narzędzie diagnostyczne zmian ogniskowych tarczycy, będąc tym samym koniecznym uzupełnieniem diagnostyki ultrasonograficznej. W kolejnych latach liczne towarzystwa naukowe na świecie opublikowały klasyfikacje szacujące ryzyko zmian ogniskowych w tarczycy, przedstawiające zróżnicowane progi wielkości oraz cechy ultrasonograficzne dla poszczególnych kategorii, w tym do wykonania biopsji.

Klasyfikacja TIRADS w swoim założeniu powinna charakteryzować się dużą czułością i tym samym pozwalać wykluczyć zmiany o niskim ryzyku z wykonywania biopsji oraz zminimalizować odsetek zmian fałszywie ujemnych. Należy jednak pamiętać, że poza czułością istotna jest możliwie wysoka swoistość, będąca przeciwieństwem do nadmiernego diagnozowania i leczenia chorych, którzy faktycznie tego nie potrzebują. Niestety nie jest możliwe uzyskanie narzędzia doskonałego, posiadającego czułość i swoistość równe 100%, dlatego istotne jest porównanie różnych klasyfikacji i zaadaptowanie najlepszej z nich do potrzeb diagnostycznych w danej populacji.

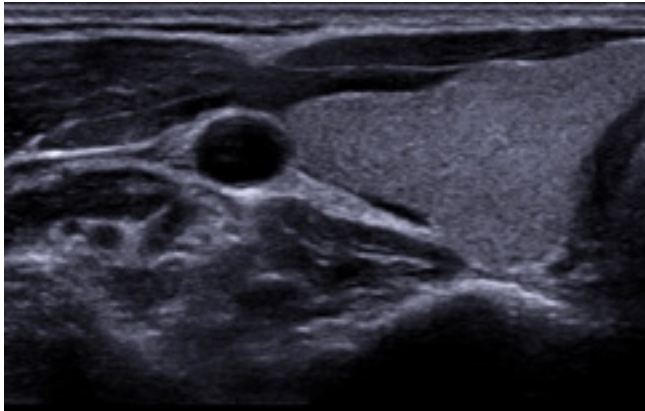
W niniejszej publikacji autorzy przedstawiają przegląd piśmiennictwa opartego na metaanalizach dotyczących najczęściej cytowanych w literaturze klasyfikacji TIRADS.

Klasyfikacja EU-TIRADS – przegląd piśmiennictwa

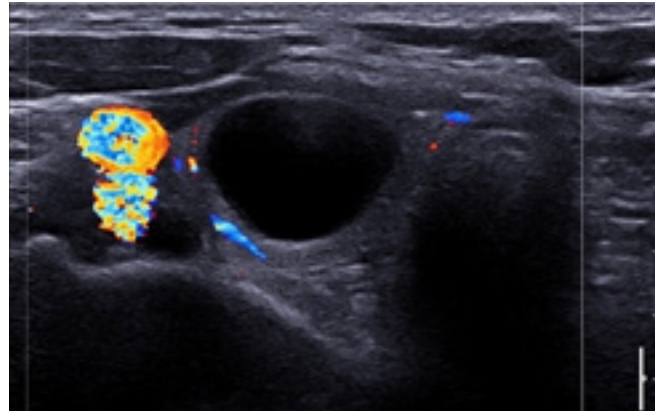
Klasyfikacja EU-TIRADS została wprowadzona w 2017 roku i przedstawiona jako wytyczne Europejskiego Towarzystwa Tyreologicznego (European Thyroid Association, ETA)⁽⁴⁾. Ten system punktacji wywodzi się z francuskiej klasyfikacji, która przed wprowadzeniem została prospektywnie zwalidowana, a jej wysoką wartość diagnostyczną potwierdzili Yoon i wsp. w badaniu 4696 guzków tarczycy (GT), w którym wykazano jej wysoką czułość i negatywną wartość predykcyjną (NPV)⁽⁵⁾. Zmiany zostały podzielone na cztery kategorie – od 2 do 5 (EU-TIRADS 1 odpowiada prawidłowej tarczycy). Opis każdej kategorii, ryzyko nowotworu oraz proponowane wskazania do biopsji przedstawiono w Tabeli 1. Opis zmian

Tab. 1. Klasyfikacja EU-TIRADS według zaleceń Europejskiego Towarzystwa Tyreologicznego (2017)

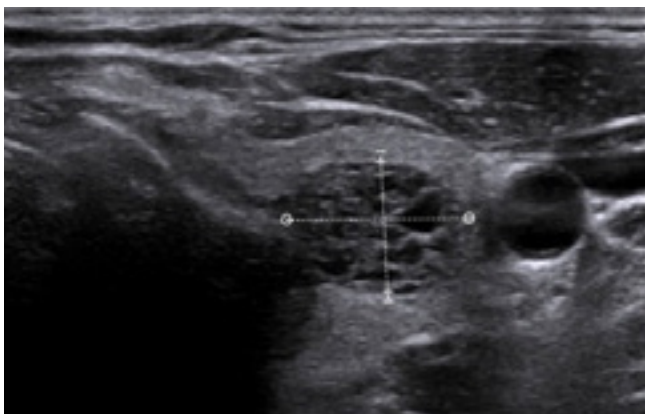
Kategoria EU-TIRADS	Rodzaj zmian	Ryzyko złośliwości	Wskazania do biopsji
1	bez zmian ogniskowych	bliskie 0%	niezalecana
2	ogniska bezechowe ogniska lito-płynowe o strukturze gąbczastej	bliskie 0%	niezalecana (wyjątek: biopsja terapeutyczna u chorych objawowych, np. opróżnienie torbieli)
3	owalny kształt równe granice/brzegi izoechogeniczne lub hiperechogeniczne bez ultrasonograficznych cech wysokiego ryzyka	2–4%	>20 mm
4	owalny kształt równe granice/brzegi nieznacznie hipoechogeniczne bez ultrasonograficznych cech wysokiego ryzyka	6–17%	>15 mm
5	głęboko hipoechogeniczne kształt inny niż owalny nierówne (nieregularne) granice/brzegi mikrozwapnienia	26–87%	>10 mm



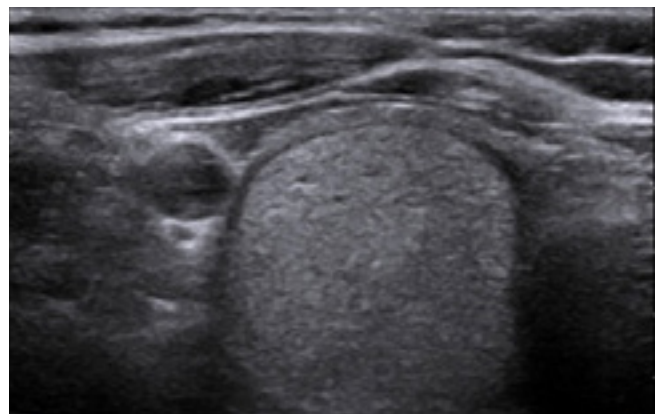
Ryc. 1. EU-TIRADS 1, prawidłowy miąższ (przekrój poprzeczny prawego płata)



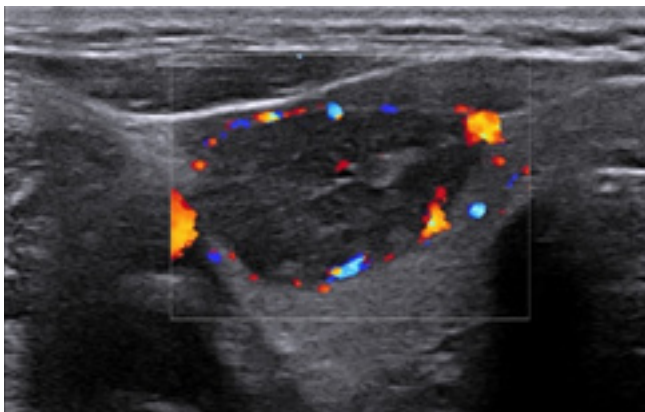
Ryc. 2. EU-TIRADS 2, prosta torbiel



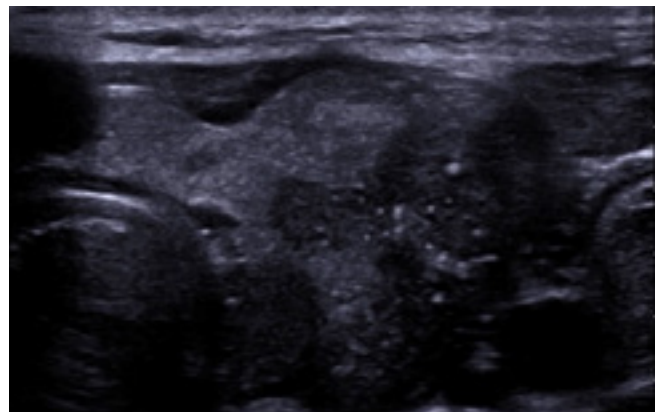
Ryc. 3. EU-TIRADS 2, zmiana gąbczasta



Ryc. 4. EU-TIRADS 3, lita, izoechogeniczna zmiana z hipoechogenicznym halo



Ryc. 5. EU-TIRADS 4, lita, hipoechogeniczna zmiana. W obrazowaniu CD unaczyniona brzeżnie oraz centralnie



Ryc. 6. EU-TIRADS 5, lita, hipoechogeniczna zmiana, z obecnością jasnych ech (mikrozwapnienia), o nierównych brzegach, nieregularnym kształcie

z każdej kategorii połączono z przykładowymi obrazami USG (Ryc. 1–6).

Rzeczywiste działanie klasyfikacji EU-TIRADS zostało przetestowane również na polskiej populacji. Po raz pierwszy została ona przeanalizowana przez Skowrońską i wsp. na grupie 52 pacjentów z obecnością 40 GT; ocenę charakteru zmian przeprowadzono w odniesieniu do badania histopatologicznego⁽⁶⁾. Odsetek raka tarczycy

wyniósł 0% w EU-TIRADS 2; 0% w EU-TIRADS 3; 5,9% w EU-TIRADS 4 i 75% w EU-TIRADS 5. Gdy jako wartość odcięcia zastosowano EU-TIRADS ≥ 4 , czułość, swoistość, wartości PPV (*positive predictive value*) i NPV wynosiły odpowiednio: 75%, 94,1%, 75% i 94,1%.

Innym badaniem mającym na celu zweryfikowanie skuteczności diagnostycznej EU-TIRADS RSS (*Risk Stratification Scale*) w warunkach polskich było wielośrodkowe badanie

opublikowane przez Dobruch-Sobczak i wsp., w którym u 428 pacjentów zidentyfikowano 842 GT (613 łagodnych, 229 złośliwych), a ich charakter potwierdzono za pomocą weryfikacji histopatologicznej⁽⁷⁾. Wskaźnik złośliwości w poszczególnych kategoriach EU-TIRADS był następujący: 0% (EU-TIRADS 2), 3,2% (EU-TIRADS 3), 11,7% (EU-TIRADS 4) i 43,5% (EU-TIRADS 5). Autorzy wykazali, że gdyby zastosować zaproponowane w wytycznych punkty odcięcia dla BACC, pominięto by 36% złośliwych GT. Wartość graniczna przy EU-TIRADS 4 miała wysoką czułość (100%), niską swoistość (25,1%), przy odcięciu dla EU-TIRADS 5 wartości wyniosły odpowiednio: 93,4% i 54,6%. Kolejnym badaniem mającym na celu ocenę skuteczności klasyfikacji EU-TIRADS w polskiej populacji było prospektywne badanie Szczepanek-Parulskiej i wsp., w którym przeanalizowano łącznie 133 GT od 88 pacjentów i uzyskano następujące wyniki: czułość 90,9%, swoistość 61,2%, PPV 69,8%, NPV 87,2% i dokładność 75,9%⁽⁸⁾. Innym badaniem przeprowadzonym w polskich warunkach było badanie przydatności klasyfikacji EU-TIRADS w diagnostyce GT z komórek Hürthle'a (HC) o niejednoznacznej cytologii. Celem jego przeprowadzenia było porównanie skuteczności diagnostycznej EU-TIRADS w dwóch grupach GT o niejednoznacznej cytologii (kategorie III–V systemu Bethesda), z komórkami Hürthle'a i bez nich. Badaniem objęto 162 HC i 378 non-HC GT z ustaloną diagnozą histopatologiczną (17,9% i 15,6% nowotworów). Autorzy doszli do wniosku, że zastosowanie klasyfikacji EU-TIRADS nie wspiera podejmowania decyzji klinicznych u pacjentów z cytologicznie niejednoznaczными GT HC, szczególnie zaklasyfikowanymi do kategorii IV systemu Bethesda do raportowania cytopatologii tarczycy (*Bethesda system for reporting thyroid cytopathology*, BSRTC)⁽⁹⁾. Stwierdzono również, że wynik EU-TIRADS koreluje z dodatnią mutacją BRAF u pacjentów z wczesnym stadium raka brodawkowatego tarczycy (*papillary thyroid carcinoma*, PTC). Wnioski wyciągnięto jednak z analizy 49 zmian, dlatego te wyniki wymagają potwierdzenia w badaniach na dużą skalę⁽¹⁰⁾.

Niedawno ukazała się praca obejmująca 80 pacjentów, popierająca ideę aktywnego nadzoru GT <1 cm zaklasyfikowanych do kategorii EU-TIRADS 5⁽¹¹⁾. Szesnastu (20,0%) z tych pacjentów przeszło operację po medianie obserwacji 57,2 miesiąca, co potwierdziło rozpoznanie PTC w 15 na 16 przypadków, a wszystkie były w remisji po 6–12 miesiącach obserwacji. Wyniki sugerują dobrą dokładność kategorii EU-TIRADS 5 w klasyfikowaniu złośliwych GT nawet o średnicy <1 cm oraz możliwość uważnego monitorowania w przypadku GT sklasyfikowanych jako EU-TIRADS 5⁽¹¹⁾.

Największym do tej pory opublikowanym wielośrodkowym badaniem podjętym w celu oceny skuteczności diagnostycznej EU-TIRADS w odniesieniu do histopatologii było badanie przeprowadzone przez Trimboli i wsp.⁽¹²⁾. To retrospektywne badanie obejmowało 1058 GT, z których 24,3% okazało się nowotworami. Częstość występowania nowotworów złośliwych wynosiła odpowiednio: 1,4%, 3,5%, 17% i 87,7% w klasach 2–5. Po połączeniu kategorii EU-TIRADS 4 i 5 osiągnięto 93% czułości i 97% NPV. Stosując zalecane kryteria dla BACC, uzyskano spadek NPV do 90,9%. Największą metaanalizę na ten temat

opublikowali Castellana i wsp.⁽¹³⁾. Włączono siedem badań oceniających 5672 GT z oceną histopatologiczną jako standard referencyjny. Częstość występowania nowotworów złośliwych w każdej kategorii EU-TIRADS wynosiła: 0,5%, 5,9%, 21,4% i 76,1%. Czułość, swoistość, PPV, NPV EU-TIRADS 5 wyniosły odpowiednio: 83,5%, 84,3%, 76,1%, 85,4%. Dalszą poprawę wyników stwierdzono po wykluczeniu dwóch badań o ograniczonej wielkości próby i niskiej częstości występowania nowotworów złośliwych w kategorii 5⁽¹³⁾. Skuteczność skali EU-TIRADS sprawdzono również w populacji pediatrycznej liczącej 24 pacjentów (31 GT), spośród których 14 przeszło tyroidektomię; wszystkie złośliwe GT zaliczono do kategorii 4 lub 5 EU-TIRADS. Czułość klasyfikacji EU-TIRADS w wykrywaniu złośliwych GT wyniosła 100%, swoistość – 25%, PPV – 44%, a NPV – 100%⁽¹⁴⁾.

Jedno z najnowszych badań Kovatchevej i wsp. miało na celu ocenę przydatności skali EU-TIRADS w kwalifikacji do BACC GT i jej zdolności do zmniejszenia liczby niepotrzebnych BACC⁽¹⁵⁾. Badanie było jednośrodkowe, prospektywne i obejmowało 741 pacjentów z 942 GT. Współczynnik złośliwości w kategoriach od 2 do 5 wynosił odpowiednio: 0, 0, 3,8 i 30,6%. Czułość, swoistość, PPV, NPV i dokładność EU-TIRADS z punktem odcięcia w kategorii 5 wyniosły odpowiednio: 91,3, 74,6, 30,6, 98,6 i 76,4%. Wydajność diagnostyczna inna niż czułość i NPV była lepsza w guzkach ≥10 mm. Autorzy zauważyli, że liczba BACC zmniejszyłaby się o 53,4%, gdyby kryteria BACC były ściśle przestrzegane. Kiedy przestrzegano wskazań do BACC, szacowana czułość, swoistość, PPV i NPV EU-TIRADS wyniosły odpowiednio: 69,9, 56,3, 16,4 i 93,8%⁽¹⁵⁾. Przydatność EU-TIRADS wykazano również w grupie 75 GT wykazujących zwiększony wychwyty radioizotopu w badaniu 18F-FDG PET/CT, które następnie oceniono w USG tarczycy. Wskaźnik raka w tej grupie wynosił 0% w EU-TIRADS 2, 2,9% w EU-TIRADS 3, 4,2% w EU-TIRADS 4 i 78,6% w EU-TIRADS 5 ($p < 0,001$). Czułość, swoistość, PPV, NPV i dokładność skali wyniosły odpowiednio: 5%, 95%, 79%, 97% i 93%. Stwierdzono zatem, że EU-TIRADS może być pomocna w przewidywaniu charakteru GT wykrytych przez 18F-FDG PET/CT⁽¹⁶⁾.

Podsumowując, skala EU-TIRADS jest prostą i wartościową RSS, charakteryzującą się bardzo dobrą czułością, łatwą do przyjęcia w praktyce klinicznej. Jej siłą jest szczególnie wysoka NPV, co umożliwia zmniejszenie liczby BACC.

ACR-TIRADS – klasyfikacja i przegląd piśmiennictwa

W 2017 roku Amerykańskie Towarzystwo Radiologiczne (*American College of Radiology*, ACR) we współpracy z Amerykańskim Stowarzyszeniem Endokrynologów Klinicznych (*American Association of Clinical Endocrinologists*, AAEC) i Amerykańskim Towarzystwem Tyreologicznym (*American Thyroid Association*, ATA) opracowało praktyczny ultrasonograficzny system służący do oceny ryzyka złośliwości guzków tarczycy nazwany ACR Thyroid Imaging, Reporting and Data

Tab. 2. Klasyfikacja ACR TI-RADS – cechy ultrasonograficzne oraz ich wartość punktowa

Cechy USG	Cechy badania USG w skali punktowej
Morfologia	torbielowata = 0 gąbczasta = 0 mieszana lito-torbielowata = 1 lita = 2 niejednorodna = 2
Echogeniczność	bezechowa = 0 izoechogeniczna = 1 hiperechogeniczna = 1 niejednorodna = 1 hipoechogeniczna = 2 głęboko hipoechogeniczna = 3
Kształt	szersza niż wyższa = 0 wyższa niż szersza = 3
Brzegi/granice	równe = 0 słabo widoczne = 0 źle zdefiniowane/nie mogą być określone = 0 nierówne (nieregularne)/płatowate = 2 naciek pozatarczycowy = 3
Obszary hiperechogeniczne	brak = 0 artefakty ogona komety = 0 makrozwapnienia = 1 obwodowe zwapnienie lub pierścieni = 2 punktowe ogniska echogeniczne (mikrozwapnienia) = 3

System (ACR-TIRADS)⁽¹⁷⁾. Podstawą koncepcji tej skali jest powszechnie i szeroko stosowany system BIRADS służący do oceny zmian ogniskowych w piersiach⁽¹⁸⁾. Nadrzędnym celem klasyfikacji ACR-TIRADS, jak również innych ultrasonograficznych RSS, jest wytypowanie wśród GT tych, które na podstawie obrazu ultrasonograficznego budzą podejrzenie nowotworowego podłoża i wymagają pogłębienia diagnostyki oraz dodatkowych procedur, najczęściej z zastosowaniem BACC. Dodatkowo klasyfikacja ta pomaga ustawić wnioski z interpretacji obrazów USG, poprawiając i usprawniając tym samym komunikację pomiędzy ultrasonografami a klinicystami, jednocześnie przyczyniając się do ograniczenia nadmiernie wykonywanych BACC.

ACR-TIRADS, w przeciwieństwie do innych skal, opiera się na unikatowej klasyfikacji punktowej. Spośród pięciu kategorii cech USG poszczególne – morfologia, echogeniczność, kształt, brzegi, mikrozwapnienia – podlegają ocenie i mają przyznawane od 0 do 3 punktów, przy czym wyższa wartość punktowa jest związana z wyższym ryzykiem wystąpienia nowotworu (Tab. 2). Łączna liczba punktów określa ryzyko złośliwości TIRADS (TR) podzielone na następujące kategorie: TR1 – łagodne (0 pkt), TR2 – niepodjęzane (2 pkt), TR3 – nieznacznie podejrzane (3 pkt), TR4 – umiarkowanie podejrzane (4–6 pkt) i TR5 – wysoce podejrzane (≥7 pkt). Łączna ocena klasy TR i maksymalnej średnicy zmiany determinuje dalsze postępowanie, które może być inwazyjne (BACC) lub obserwację. Szacowane ryzyko nowotworu złośliwego w poszczególnych grupach wynosi: <2% dla TR1 i TR2, <5% dla TR3, w zakresie od 5,1% do 20% dla TR4 i >20% dla TR5. Ponadto ACR-TIRADS stanowi część nowego, interaktywnego internetowego algorytmu,

zwanego TNAPP (Thyroid Nodule App), który został przedstawiony podczas corocznej konferencji American Association for Clinical Endocrinology w 2021 roku⁽¹⁹⁾. Innowacyjność TNAPP polega na połączeniu czynników klinicznych, cech USG oraz BACC i dostarczeniu jasnych sugestii dotyczących dalszego postępowania w GT.

Wartość diagnostyczna ACR-TIRADS została potwierdzona w wielu ostatnio przeprowadzonych badaniach. W metaanalizie opublikowanej przez Li i wsp., obejmującej 16 badań i oceniającej 21 882 GT, autorzy wykazali, że metaanalityczna całościowa czułość i swoistość w stratyfikacji ryzyka wynosiła: 89% (95% CI, 81–93) i 70% (95% CI, 60–78)⁽²⁰⁾. Diagnostyczny iloraz szans (*diagnostic odds ratio*, DOR) wyniósł 18,46 (95% CI 9,77–34,88). Wśród tych badań 10/16 artykułów porównało bezpośrednią przydatność diagnostyczną ACR z wytycznymi ATA, a 6 badań porównywało ACR z koreańskim TIRADS, obserwując podobną czułość: 83% vs 87% i 85% vs 91%, ale wyższą swoistość ACR w porównaniu z dwiema pozostałymi (65% vs 50%, 57% vs 24%).

Podsumowując, ACR-TIRADS jest uznanym, wartościowym narzędziem stosowanym w ocenie ryzyka złośliwości GT, które dostarcza szczegółowej, rozległej charakterystyki zmian z jednoznacznymi zaleceniami dotyczącymi postępowania, co pozwala uniknąć niepotrzebnej BACC. Główną wadą tego systemu w zastosowaniu w codziennej praktyce klinicznej jest dość duża czasochłonność w jej wykorzystywaniu, w szczególności dla początkujących lekarzy.

Klasyfikacja: K-TIRADS – przegląd piśmiennictwa

W 2016 roku Korean Society of Thyroid Radiology (KSThR) przedstawiło klasyfikację K-TIRADS oraz rekomendacje, opisując terminologię i symptomatologię obrazu ultrasonograficznego zmian ogniskowych tarczycy⁽²¹⁾. Opisane cechy obrazu USG zostały wykorzystane przy tworzeniu klasyfikacji K-TIRADS oraz zaleceń względem poszczególnych kategorii. Poza obrazem B-mode autorzy wyszczególnili również typy unaczynienia zmian ogniskowych oraz potencjalną rolę sonoelastografii w diagnostyce różnicowej GT, nie umieszczając ich jednak w samej klasyfikacji ze względu na rozbieżne wyniki prezentowane w literaturze. Sama klasyfikacja opiera się na ultrasonograficznych cechach obrazu B-mode. Za cechy podejrzane uznano: mikrozwapnienia, pionowy kształt zmiany, brzegi spikularne/mikrolobularne⁽²¹⁾. K-TIRADS 1 odpowiada zdrowemu mięszowi tarczycy bez obecnych zmian ogniskowych. K-TIRADS 2 należy przydzielić torbielom, zmianom częściowo torbielowatym z obecnymi artefaktami ogonów komety oraz guzkom gąbczastym. W przypadku zmian częściowo torbielowatych izo-, hiperechogenicznych lub o mieszanej echogeniczności (izo-/hiper-) bez którejkolwiek z ultrasonograficznych cech podejrzanych przydzieliła się kategorię „3”, podczas gdy zmianom z obecną dowolną cechą podejrzaną przydzieliła się kategorię „4”. Podobnie kategorię 4 przydzieliła się zmianom litym, hipoechogenicznym bez obecności cech podejrzanych. K-TIRADS 5 dotyczy zmian

litych, hipoechogenicznych z obecną co najmniej jedną cechą podejrzaną. Dalsze postępowanie ze zmianami ogniskowych jest uzależnione od ich wielkości (dotyczy największego wymiaru)⁽²¹⁾:

- kategoria 2:
 - zmiany gąbczaste BACC ≥ 2 cm (ryzyko złośliwości, RZ <3%),
 - zmiany częściowo torbielowate z obecnością artefaktów ogonów komety, BACC niezalecana (RZ <1%),
- kategoria 3: BACC $\geq 1,5$ cm (RZ 3–15%),
- kategoria 4: BACC ≥ 1 cm (RZ 15–30%),
 - lita, hipoechogeniczna zmiana ogniskowa bez cech podejrzanых,
 - zmiana częściowo torbielowata lub izo-/hiperechogeniczna z obecną dowolną cechą podejrzaną,
- kategoria 5: BACC ≥ 1 cm lub wybiórczo zmiany >5 mm (RZ >60%).

Klasyfikacja K-TIRADS w porównaniu z klasyfikacją ACR-TIRADS wydaje się bardziej intuicyjna, jednakże należy pamiętać, że klasyfikacja ACR-TIRADS pozwala (również wymusza) na bardziej obiektywną ocenę (analizę) każdej ze zmian z osobna (w przeciwieństwie do K- i EU-TIRADS). Poza tym należy zwrócić uwagę, że zmiany ogniskowe z cechami podejrzanymi oraz o budowie mieszanej lub litej i izo-/hiperechogeniczne w klasyfikacji K-TIRADS znajdują się w kategorii 4⁽²¹⁾, podobnie jak w ACR-TIRADS⁽¹⁷⁾, podczas gdy w klasyfikacji EU-TIRADS są przydzielane do kategorii 5⁽⁴⁾. K-TIRADS, podobnie jak EU-TIRADS, nie traktują makrozwapnień, zwapnień obrączkowatych oraz nacieku poza torebkę tarczycy jako cechy o wyższym ryzyku w przeciwieństwie do ACR-TIRADS, pomijając tym samym dość istotne cechy GT uznane w literaturze za podejrzanе. Kolejną rozbieżność dla tych klasyfikacji jest zauważalna w progach odcięcia wielkości guzków kierowanych do BACC. Autorzy K-TIRADS zalecają weryfikację cytologiczną w kategoriach 4. i 5. w przypadku zmian ≥ 10 mm, zaznaczając możliwość wybiórczej BACC w kategorii 5 dla zmian >5 mm. W przypadku EU-TIRADS i ACR-TIRADS dla kategorii 4 BACC jest zalecana w przypadku, gdy zmiana jest wielkości >15 mm oraz >10 mm w kategorii 5^(4,17,21). W metaanalizie Kim i wsp. przedstawiają porównanie możliwości diagnostycznych klasyfikacji K, ACR- oraz EU-TIRADS względem punktów odcięcia dla kategorii 5 (Tab. 3) i 4 lub 5 (Tab. 4) – nie stwierdzono istotnych statystycznie różnic pomiędzy nimi. Zauważalny był jednak trend wyższych wartości czułości dla EU-TIRADS i wyższej swoistości dla K-TIRADS, a wyniki czułości i swoistości przedstawiono w Tab. 3 i Tab. 4⁽²²⁾. W drugiej metaanalizie, porównującej możliwości diagnostyczne tych klasyfikacji, uzyskano podobny wniosek końcowy – wyniki dla czułości i swoistości dla tych klasyfikacji względem kategorii 5 (Tab. 5) oraz 4 lub 5 (Tab. 6) są zbliżone⁽²³⁾.

Dyskusja na temat klasyfikacji TIRADS powinna również obejmować temat biopsji cienkoigłowej, a dokładniej rzecz ujmując – odsetka niepotrzebnych procedur generowanych

Tab. 3. Porównanie parametrów diagnostycznych klasyfikacji K-, ACR- i EU-TIRADS względem punktu odcięcia dla kategorii 5⁽²²⁾

Kategoria 5	Czułość (%)	Swoistość (%)
K-TIRADS	64	93
ACR-TIRADS	70	89
EU-TIRADS	78	89

Tab. 4. Porównanie parametrów diagnostycznych klasyfikacji K-, ACR- i EU-TIRADS względem punktu odcięcia dla kategorii 4 lub 5⁽²²⁾

Kategoria 4 lub 5	Czułość (%)	Swoistość (%)
K-TIRADS	92	61
ACR-TIRADS	95	49
EU-TIRADS	96	48

Tab. 5. Porównanie parametrów diagnostycznych klasyfikacji K-, ACR- i EU-TIRADS względem punktu odcięcia dla kategorii 5⁽²³⁾

Kategoria 5	Czułość (%)	Swoistość (%)
K-TIRADS	55	95
ACR-TIRADS	66	91
EU-TIRADS	82	90

Tab. 6. Porównanie parametrów diagnostycznych klasyfikacji K-, ACR- i EU-TIRADS względem punktu odcięcia dla kategorii 4 lub 5⁽²³⁾

Kategoria 4 lub 5	Czułość (%)	Swoistość (%)
K-TIRADS	89	64
ACR-TIRADS	95	55
EU-TIRADS	96	52

przez poszczególne klasyfikacje. Powyższy aspekt został szerzej omówiony w kolejnej metaanalizie⁽²⁴⁾. W tej pracy autorzy wskazują, że klasyfikacja K-TIRADS charakteryzuje się dużym odsetkiem niepotrzebnych biopsji – na poziomie 55%. Najniższy odsetek charakteryzuje klasyfikację ACR-TIRADS (25%), istotnie niższy niż dla K-TIRADS ($p < 0,05$) i niższy w porównaniu z EU-TIRADS (38%; $p = 0,087$). Oczywiście jest to, że ten odsetek jest ściśle związany zarówno ze stratyfikacją zmian ogniskowych (przy porządkowaniu do poszczególnych kategorii), jak i punktem odcięcia wielkości guzka. Tym samym zapożyczenie punktów odcięcia z klasyfikacji ACR-TIRADS i zastosowanie ich w klasyfikacjach K-TIRADS oraz EU-TIRADS zmniejszy odsetek niepotrzebnych BACC⁽²⁵⁾. Należy jednak pamiętać, że redukcja tzw. niepotrzebnych biopsji nie musi być właściwym rozwiązaniem. Może to pociągać za sobą zwiększenie odsetka nierozpoznanych nowotworów, co może negatywnie wpływać na krzywą przeżycia chorych z rakiem tarczycy. Na podstawie samych metaanaliz wyciągnięcie ostatecznego wniosku na temat odsetka niepotrzebnych biopsji jest trudne. Czynnikiem ułatwiającym decyzję może być element ekonomiczny, tj. odsetek niepotrzebnych biopsji i tym samym możliwy odsetek wyników fałszywie dodatnich wraz z ich konsekwencjami, jaki badacze są w stanie zaakceptować.

Podsumowanie metaanaliz badań uwzględniających ocenę porównawczą *head to head* zastosowania klasyfikacji TIRADS w szacowaniu potencjału złośliwości zmian ogniskowych gruczołu tarczowego – przegląd piśmiennictwa

Przedmiotem analizy jest ocena zastosowania pięciu głównych klasyfikacji: ACR-TIRADS (American College of Radiology guidelines), ATA (American Thyroid Association guidelines), Kwak-TIRADS, K-TIRADS (Korean Thyroid Association/Korean Society of Thyroid Radiology (KTA/KSThR) guidelines) (K-TIRADS jest rozwinięciem klasyfikacji Kwak-TIRADS) oraz EU-TIRADS (European Thyroid Association (ETA) guidelines) w szacowaniu potencjału złośliwości zmian ogniskowych tarczycy. Na przełomie ostatnich kilku lat opublikowano wiele badań porównujących zastosowanie klasyfikacji TIRADS w ocenie ryzyka złośliwości zmian ogniskowych tarczycy, których wybrane elementy przedstawiono w poniższym streszczeniu.

Yang i wsp. przeprowadzili metaanalizę 19 badań oceniających użyteczność kliniczną stosowania 5 wyżej wskazanych klasyfikacji. Metaanaliza obejmuje wyniki prac badawczych przedstawionych w artykułach, które ukazały się w latach 2015–2020. Ocenie poddano od 100 do 4696 zmian ogniskowych gruczołu tarczowego (łącznie 24 325 zmian) diagnozowanych u od 92 do 4585 pacjentów. Wszystkie zmiany ogniskowe tarczycy zostały poddane ocenie cyto- i/lub histopatologicznej. Dwanaście spośród poddanych analizie badań uwzględniało ocenę ACR-TIRADS, 10 – klasyfikację wg ATA, 6 – Kwak-TIRADS oraz 4 – klasyfikację EU-TIRADS, a 4 – badania K-TIRADS. Na podstawie przeprowadzonej analizy statystycznej uzyskano wysoką czułość poszczególnych klasyfikacji w zakresie 0,84–0,96 (95% CI),

natomiast swoistość przedstawiała się kolejno dla: ACR-TIRADS, ATA, Kwak-TIRADS, KTA i EU-TIRADS, odpowiednio: 0,68, 0,44, 0,62, 0,47 i 0,61 (95% CI) – Tab. 7. Dla wszystkich klasyfikacji uzyskano wartość pola pod krzywą ROC (AUC) >0,8, z rozkładem wartości dla powyższych kwalifikacji, odpowiednio: 0,8553, 0,8976, 0,9101, 0,9022 i 0,8810, co przemawia za bardzo dobrą dokładnością diagnostyczną ich zastosowania⁽²⁶⁾ (Tab. 7).

Natomiast w metodzie porównawczej *head to head* uzyskano wartości względnego diagnostycznego ilorazu szans RDOR w zakresie: 1,57 (ACR vs ATA), 1,37 (ACR vs EU), 1,8 (ACR vs Kwak), 1,74 (ACR vs K), co wyróżnia ACR-TIRADS w tej analizie (Tab. 8). Autorzy pracy podkreślili ponadto różnice w systemie klasyfikacyjnym TIRADS. ACR i Kwak-TIRADS oparto na klasyfikacji punktowej, natomiast pozostałe kategoryzują zmiany ogniskowe w oparciu o wzorce sonograficzne, co w praktyce klinicznej wydaje się postępowaniem bardziej intuicyjnym, natomiast charakteryzują się mniejszą dokładnością testu. Przykładowo kategoria 5 lub 4 wg EU-TIRADS może odpowiadać ACR-TIRADS TR4/3 lub K-TIRADS TR4/3, natomiast zmiany ogniskowe zakwalifikowane do K-TIRADS TR3 i EU-TIRADS kategorii 3 (niskie ryzyko złośliwości) mogą być skategoryzowane jako ACR-TIRADS TR2, co oznacza brak podejrzenia złośliwości. Powyższe różnice w kryteriach kwalifikacyjnych, uwzględniające również wielkość zmiany, prowadzi do powstania różnic w swoistości stosowanych metod. Na podstawie przeprowadzonej metaanalizy dla ACR-TIRADS odnotowano najwyższą dokładność diagnostyczną w szacowaniu ryzyka złośliwości i związane z tym ograniczenie wskazań do wykonywania BACC⁽²⁶⁾. Autorzy pracy wskazują na ograniczenia przeprowadzonej metaanalizy, do których należy brak weryfikacji histopatologicznej w przypadku każdej analizowanej zmiany ogniskowej.

Tab. 7. Czulość, swoistość, LR(+), LR(-), DOR, AUC dla poszczególnych kwalifikacji TIRADS

Klasyfikacja TIRADS	Liczba badań	Czułość (95% CI)	Swoistość (95% CI)	LR(+) (95% CI)	LR(-) (95% CI)	DOR (95% CI)	AUC
ACR-TIRADS	13	0,85 (0,84–0,86)	0,68 (0,6–0,69)	2,98 (2,37–3,75)	0,22 (0,16–0,29)	15,23	0,8553
EU-TIRADS	4	0,85 (0,83–0,87)	0,61 (0,59–0,62)	2,84 (1,43–5,64)	0,21 (0,13–0,34)	13,18	0,8810
K-TIRADS	4	0,85 (0,83–0,86)	0,47 (0,46–0,48)	2,60 (1,2–5,57)	0,18 (0,08–0,39)	14,57	0,9022
Kwak-TIRADS	6	0,94 (0,94–0,95)	0,62 (0,6–0,63)	3,23 (0,90–11,61)	0,08 (0,04–0,16)	43,15	0,9101

CI (confidence interval) – przedział ufności; LR(+) (positive likelihood ratio) – pozytywny iloraz wiarygodności; LR(-) (negative likelihood ratio) – negatywny iloraz wiarygodności; DOR (diagnostic odds ratio) – diagnostyczny iloraz szans; AUC (area under curve) – pole pod wykresem krzywej ROC

Tab. 8. Porównanie *head to head* względnego diagnostycznego ilorazu szans (RDOR) z przedziałem ufności CI 95%

	ACR-TIRADS	EU-TIRADS	K-TIRADS	Kwak-TIRADS
ACR-TIRADS	–	0,7308 (0,3000–1,7803)	0,5734 (0,2759–1,1919)	0,5564 (0,2552–1,2131)
EU-TIRADS	1,3683 (0,5617–3,3332)	–	0,7846 (0,3075–2,0020)	0,7614 (0,2498–2,3208)
K-TIRADS	1,7439 (0,8390–3,6247)	1,2745 (0,4995–3,2518)	–	0,9703 (0,3697–2,5466)
Kwak-TIRADS	1,7972 (0,8243–3,9183)	1,3138 (0,4309–4,0035)	1,0306 (0,3927–2,7048)	–

RDOR (relative diagnostic odds ratio) – względny diagnostyczny iloraz szans; CI (confidence interval) – przedział ufności

Kolejnym ograniczeniem badania jest różnica liczbowa zakwalifikowanych do analizy zmian o potencjale złośliwym, poza tym nie ma wystarczających danych związanych z porównawczą analizą z zastosowaniem K-TIRADS oraz EU-TIRADS. Na uwagę zwraca również ograniczenie wynikające z niezgodności wyników badań ultrasonograficznych uzyskanych przez różnych badających i przez tego samego badającego (*the inter-observer and intra-observer variability*)⁽²⁷⁾. Wnioski płynące z przedstawionej metaanalizy podkreślają użyteczność kliniczną oraz potwierdzają zasadność rekomendacji zastosowania klasyfikacji TIRADS w szacowaniu potencjału złośliwości zmian ogniskowych tarczycy, z podkreśleniem najwyższej dokładności metody dla ACR-TIRADS.

Zbieżne wyniki dotyczące korzyści zastosowania klasyfikacji ACR-TIRADS uzyskali Castellana i wsp., którzy przeprowadzili metaanalizę 12 badań (z czego 6 objętych poprzednią metaanalizą) uwzględniających 5 klasyfikacji (AAACE/ACE/AME, ACR-TIRADS, ATA, EU-TIRADS i K-TIRADS) w celu oceny ich zastosowania do typowania GT do BACC⁽²⁸⁾. Analizie poddano łącznie 18 750 GT (w tym 4378 zmian złośliwych zweryfikowanych histopatologicznie oraz 14 372 zmian łagodnych zweryfikowanych cyto- i/lub histopatologicznie). Uzyskano wartość wskaźnika DOR w zakresie od 2,2 do 4,9, z najwyższą wartością RDOR w ocenie porównawczej *head to head* dla ACR-TIRADS vs ATA ($p = 0,02$) oraz vs K-TIRADS ($p = 0,002$), co było spowodowane najwyższym wskaźnikiem ilorazu wiarygodności dla wyniku pozytywnego (LR+ (95% CI)). Powyższe wyniki potwierdzają przewagę zastosowania dla ACR-TIRADS w typowaniu zmian do BACC, a co za tym idzie – ograniczeniu liczby niepotrzebnie wykonywanych zabiegów.

W kolejnej pracy, przeprowadzonej przez Kim i wsp., przeanalizowano łącznie 34 badania (poddające analizie łącznie 37 585 zmian ogniskowych tarczycy), w których oceniano użyteczność diagnostyczną zastosowania TIRADS w szacowaniu ryzyka złośliwości zmian ogniskowych tarczycy poprzez ich klasyfikację do kategorii 4 i 5 TIRADS (TR4, TR5) z uwzględnieniem 3 klasyfikatorów: EU-TIRADS, ACR-TIRADS oraz K-TIRADS⁽²²⁾. Dla TR5 odnotowano najwyższą czułość EU-TIRADS: 78% (95% CI, 64–88%), następnie ACR-TIRADS: 70% (95% CI, 61–79%) i K-TIRADS: 64% (95% CI, 58–70%). Najwyższą swoistość w stosunku do zastosowania TR5 zaobserwowano dla K-TIRADS: 93% (95% CI, 91–95%), ACR-TIRADS: 89% (95% CI, 85–92%) i EU-TIRADS: 89% (95% CI, 77–95%). Kategoryzacja do TR4/5 we wszystkich 3 klasyfikacjach charakteryzuje czułość >90%, natomiast najwyższą swoistość odnotowano dla K-TIRADS: 61% (95% CI, 50–72%), chociaż nie zaobserwowano istotnej statystycznie różnicy pomiędzy klasyfikacjami. Wyniki metaanalizy wskazują jednak na trend w kierunku najwyższej czułości dla zastosowania EU-TIRADS oraz najwyższej swoistości dla K-TIRADS jako narzędzi diagnostycznych w szacowaniu ryzyka złośliwości zmian ogniskowych tarczycy.

W kolejnej metaanalizie 29 prac oceniających łącznie 33 748 zmian ogniskowych tarczycy Kim i wsp. odnotowali

porównywalną dokładność diagnostyczną 4 przedstawionych klasyfikacji TIRADS w szacowaniu ryzyka złośliwości. Czułość i swoistość metody uwzględniano w stosunku do ACR-TIRADS, wynoszącą odpowiednio: 66% i 91% dla TR5 oraz 95% i 55% dla TR4/5, w klasyfikacji K-TIRADS: 55% i 95% dla TR5 oraz 89% i 64% dla TR4/5, a także w systemie EU-TIRADS: 82% i 90% dla TR5 oraz 96% i 52% dla TR4/5⁽²³⁾. Nie odnotowano istotnej przewagi w przypadku żadnej z klasyfikacji jako czynnika prognostycznego ryzyka złośliwości zmian ogniskowych. Zwrócono uwagę na czynniki wpływające na ograniczenia przeprowadzonej metaanalizy – zalicza się do nich różne populacje, na których przeprowadzono badanie, kryteria włączenia, proporcję liczbową pacjentów różnej płci oraz liczbę zmian złośliwych poddanych analizie.

Kim i wsp. w dokonanej metaanalizie 8 badań z analizą 13 092 zmian ogniskowych tarczycy odnotowali najniższy wskaźnik niepotrzebnie wykonywanych BACC w stosunku do klasyfikacji ACR-TIRADS, wynoszący 25% (95% CI, 22–29%), istotnie statystycznie niższy w porównaniu z ATA: 51% (95% CI, 44–58% ($p < 0,001$)) oraz K-TIRADS: 55% (95% CI, 42–67% ($p < 0,001$)). Nie odnotowano natomiast istotnej zależności w porównaniu z EU-TIRADS: 38% (95% CI, 16–66%, ($p = 0,087$))⁽²⁴⁾. Powyższa analiza dowodzi największej korzyści z zastosowania klasyfikacji ACR-TIRADS oraz EU-TIRADS w typowaniu zmian do BACC, co ma przełożenie na minimalizowanie liczby niepotrzebnie wykonywanych inwazyjnych procedur medycznych.

Podsumowanie

Przedstawione wyniki metaanaliz dowodzą użyteczności diagnostycznej zastosowania klasyfikacji TIRADS w szacowaniu ryzyka złośliwości zmian ogniskowych tarczycy oraz potwierdzają zasadność rekomendacji ich użycia w praktyce klinicznej. Przeprowadzona analiza statystyczna *head to head* dostępnych danych przemawia za najwyższą korzyścią stosowania klasyfikacji ACR-TIRADS, co ma również przełożenie na minimalizowanie liczby niepotrzebnie wykonywanych BACC. W celu wyciągnięcia dalej idących wniosków konieczne jest jednak przeprowadzenie większej liczby badań z zastosowaniem poszczególnych klasyfikacji TIRADS (w tym EU-TIRADS, K-TIRADS), uwzględniając przedstawione powyżej ograniczenia związane z heterogennością grup poddanych analizie badawczej. Zaletą skali EU-TIRADS jest to, że zawiera tylko najważniejsze cechy ultrasonograficzne, dzięki czemu jej zastosowanie nie jest czasochłonne, a skala jest łatwa do zastosowania w praktyce klinicznej. Wykorzystanie systemu ułatwiłoby raportowanie wyniku badania USG tarczycy i mogłoby usprawnić kwalifikację GT do BACC.

Konflikt interesów

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

Wkład autorów

Oryginalna koncepcja pracy: KD-S. Pisanie manuskryptu: KD-S, AZ, BM, AS-S, ES-P, KZ, AŻ. Analiza i interpretacja danych: KD-S, AZ, MD, BM, AS-S, ES-P, KZ, AŻ. Ostateczna akceptacja manuskryptu: KD-S, AZ, MD, AL, BM, MR, AS-S, ES-P, KZ, AŻ. Gromadzenie, rejestrowanie i/lub kompilacja danych: KD-S, AZ, BM, AS-S, ES-P, KZ, AŻ. Krytyczna recenzja rękopisu: MD, AL, MR.

Piśmiennictwo

1. Ginat DT, Butani D, Giampoli EJ, Patel N, Dogra V: Pearls and pitfalls of thyroid nodule sonography and fine-needle aspiration. *Ultrasound Q* 2010; 26: 171–178.
2. Remonti LR, Kramer CK, Leitão 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.
3. 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.
4. 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.
5. Yoon JH, Han K, Kim E-K, Moon HJ, Kwak JY: Diagnosis and management of small thyroid nodules: a comparative study with six guidelines for thyroid nodules. *Radiology* 2017; 283: 560–569.
6. Skowrońska A, Milczarek-Banach J, Wiechno W, Chudziński W, Żach M, Mazurkiewicz M *et al.*: Accuracy of the European Thyroid Imaging Reporting and Data System (EU-TIRADS) in the valuation of thyroid nodule malignancy in reference to the post-surgery histological results. *Pol J Radiol* 2018; 83: e579–586.
7. Dobruch-Sobczak K, Adamczewski Z, Szczepanek-Parulska E, Mígda B, Woliński K, Krauze A *et al.*: Histopathological verification of the diagnostic performance of the EU-TIRADS classification of thyroid nodules—results of a multicenter study performed in a previously iodine-deficient region. *J Clin Med* 2019; 8: E1781.
8. Szczepanek-Parulska E, Wolinski K, Dobruch-Sobczak K, Antosik P, Ostalowska A, Krauze A *et al.*: S-Detect Software vs. EU-TIRADS Classification: a dual-center validation of diagnostic performance in differentiation of thyroid nodules. *J Clin Med* 2020; 9: E2495.
9. Słowińska-Klencka D, Wysocka-Konieczna K, Klencki M, Popowicz B: Usability of EU-TIRADS in the diagnostics of Hürthle cell thyroid nodules with equivocal cytology. *J Clin Med* 2020; 9: E3410.
10. Skubisz K, Januszkiewicz-Caulier J, Cybula P, Bakula-Zalewska E, Goryca K, Paziewska A *et al.*: Higher EU-TIRADS-score correlated with BRAF V600E positivity in the early stage of papillary thyroid carcinoma. *J Clin Med* 2021; 10: 2304.
11. Rozenbaum A, Buffet C, Bigorgne C, Royer B, Rouxel A, Bienvenu M *et al.*: Outcomes of active surveillance of EU-TIRADS 5 thyroid nodules. *Eur J Endocrinol* 2021; 184: 677–686.
12. Trimboli P, Ngu R, Royer B, Giovanella L, Bigorgne C, Simo R *et al.*: A multicentre validation study for the EU-TIRADS using histological diagnosis as a gold standard. *Clin Endocrinol* 2019; 91: 340–347.
13. Castellana M, Grani G, Radzina M, Guerra V, Giovanella L, Deandrea M *et al.*: Performance of EU-TIRADS in malignancy risk stratification of thyroid nodules: a meta-analysis. *Eur J Endocrinol* 2020; 183: 255–264.
14. Yeste Fernández D, Vega Amenabar E, Coma Muñoz A, Arciniegas Vallejo L, Clemente León M, Planes-Conangla M *et al.*: Ultrasound criteria (EU-TIRADS) to identify thyroid nodule malignancy risk in adolescents. Correlation with cyto-histological findings. *Endocrinol Diabetes Nutr (Engl Ed)* 2021; S2530-0164(21)00078-1.
15. Kovatcheva RD, Shinkov AD, Dimitrova ID, Ivanova RB, Vidinov KN, Ivanova RS: Evaluation of the diagnostic performance of EU-TIRADS in discriminating benign from malignant thyroid nodules: a prospective study in one referral center. *Eur Thyroid J* 2021; 9: 304–312.
16. Trimboli P, Paone G, Treglia G, Virili C, Ruberto T, Ceriani L *et al.*: Fine-needle aspiration in all thyroid incidentalomas at 18 F-FDG PET/CT: Can EU-TIRADS revise the dogma? *Clin Endocrinol* 2018; 89: 642–648.
17. 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.
18. Vanel D: The American College of Radiology (ACR) Breast Imaging and Reporting Data System (BI-RADS): a step towards a universal radiological language? *Eur J Radiol* 2007; 61: 183.
19. Garber JR, Papini E, Frasoldati A, Lupo MA, Harrell RM, Parangi S *et al.*: American Association of Clinical Endocrinology and Associazione Medici Endocrinologi Thyroid Nodule Algorithmic Tool. *Endocr Pract* 2021; 27: 649–660.
20. Li W, Wang Y, Wen J, Zhang L, Sun Y: Diagnostic performance of American College of Radiology TI-RADS: a systematic review and meta-analysis. *AJR Am J Roentgenol* 2021; 216: 38–47.
21. Shin JH, Baek JH, Chung J, Ha EJ, Kim J-H, Lee YH *et al.*: Ultrasonography diagnosis and imaging-based management of thyroid nodules: Revised Korean Society of Thyroid Radiology Consensus Statement and Recommendations. *Korean J Radiol* 2016; 17: 370–395.
22. Kim PH, Chung SR, Choi SH, Kim KW: Accuracy of thyroid imaging reporting and data system category 4 or 5 for diagnosing malignancy: a systematic review and meta-analysis. *Eur Radiol* 2020; 30: 5611–5624.
23. Kim PH, Suh CH, Baek JH, Chung SR, Choi YJ, Lee JH: Diagnostic performance of four ultrasound risk stratification systems: a systematic review and meta-analysis. *Thyroid* 2020; 30: 1159–1168.
24. Kim PH, Suh CH, Baek JH, Chung SR, Choi YJ, Lee JH: Unnecessary thyroid nodule biopsy rates under four ultrasound risk stratification systems: a systematic review and meta-analysis. *Eur Radiol* 2021; 31: 2877–2885.
25. Ha SM, Baek JH, Na DG, Suh CH, Chung SR, Choi YJ *et al.*: Diagnostic performance of practice guidelines for thyroid nodules: thyroid nodule size versus biopsy rates. *Radiology* 2019; 291: 92–99.
26. Yang R, Zou X, Zeng H, Zhao Y, Ma X: Comparison of diagnostic performance of five different ultrasound TI-RADS classification guidelines for thyroid nodules. *Front Oncol* 2020; 10: 598225.
27. Liu H, Ma A-L, Zhou Y-S, Yang D-H, Ruan J-L, Liu X-D *et al.*: Variability in the interpretation of grey-scale ultrasound features in assessing thyroid nodules: a systematic review and meta-analysis. *Eur J Radiol* 2020; 129: 109050.
28. Castellana M, Castellana C, Treglia G, Giorgino F, Giovanella L, Russ G *et al.*: Performance of five ultrasound risk stratification systems in selecting thyroid nodules for FNA. *J Clin Endocrinol Metab* 2020; 105: dgz170.