

Nabil Abdalla

Promotor: Prof. dr hab. n. med. Włodzimierz Sawicki

Ocena przydatności diagnostycznej wybranych matematycznych modeli i systemów punktacji w przedoperacyjnym różnicowaniu guzów przydatków

STRESZCZENIE

Wstęp

Guzy przydatków stanowią różnorodną grupę, obejmującą zarówno zmiany łagodne jak i złośliwe. Wybór właściwego postępowania uzależniony jest od danych z wywiadu, badania klinicznego, markerów nowotworowych, a także wyników badań obrazowych, zwłaszcza ultrasonografii i tomografii komputerowej. W oparciu o cechy kliniczne guza, parametry ultrasonograficzne i stężenie markerów nowotworowych, opracowano liczne modele matematyczne i systemy punktacji, zwiększające wartość diagnostyczną stosowanych metod.

Cel

Celem pracy była ocena wartości diagnostycznej wybranych modeli matematycznych oraz systemów punktacji w przedoperacyjnym różnicowaniu guzów przydatków. Aby zrealizować powyższe dokonano:

1. Porównania wartości diagnostycznej algorytmu oceny ryzyka złośliwości nowotworów jajnika ROMA z poszczególnymi markerami CA125 i HE4.
2. Porównania skuteczności zmodyfikowanych wskaźników mRMI 1-4 w których CA125 zastąpiono HE4 z klasycznymi wskaźnikami RMI 1-4, w różnicowaniu złośliwych i niezłośliwych guzów przydatków.

Material i Metody

Badanie prospektywne podzielono na dwa etapy. Materiał stanowił 312 pacjentek, z tego 302 kobiety było objętych analizą w pierwszym etapie, a 10 dodatkowych w etapie drugim.

W pierwszym etapie analizie poddano pacjentki z guzami przydatków hospitalizowane i operowane od października 2012 do kwietnia 2015 roku, w Katedrze i Klinice Położnictwa,

Chorób Kobięcych i Ginekologii Onkologicznej, II Wydziału Lekarskiego, Warszawskiego Uniwersytetu Medycznego, u których zastosowano algorytm ROMA.

Kryteria włączenia do badania obejmowały wiek powyżej 18 lat, dokonanie pomiaru CA125 i HE4 nie później niż do pięciu dni od operacji, uzyskanie oceny histopatologicznej materiału pooperacyjnego oraz zgodę pacjentki.

Kryteria wykluczające z analizy stanowiły: ciąża, choroby nerek, choroby nowotworowe w wywiadzie, chemio i/lub radioterapia, mięśniaki macicy o średnicy powyżej 5 cm oraz brak wyniku badania histopatologicznego zmian przydatków.

Poziom HE4 i CA125 oceniano metodą elektrochemiluminescencji, z tej samej próbki, na jednym aparacie (Cobas 8000-e602). Do wyliczenia algorytmu ROMA wykorzystano dwa modele regresji logistycznej opisane przez Moore i współautorów. Wartość predykcyjna PI wyliczana była osobno dla kobiet przed i po menopauzie. W grupie przed menopauzą użyto następującej formuły: $PI = -12,0 + 2,38 * \ln(HE4) + 0,0626 * \ln(CA125)$, natomiast po menopauzie: $PI = -8,09 + 1,04 * \ln(HE4) + 0,732 * \ln(CA125)$. W obu wzorach \ln oznacza naturalny logarytm.

Do wyliczenia ryzyka złośliwości opartego na algorytmie ROMA (%) zastosowano formułę: $\% = \exp(PI) / [1 / \exp(PI)] * 100$. Za wartość graniczną dla CA125 przyjęto 35 U/mL, dla HE4 70 i 140 pmol/L, natomiast dla algorytmu ROMA przed i po menopauzie odpowiednio 11,4 % i 29,9%. Ostateczna ocena guzów przydatków oparta była na pooperacyjnym wyniku badania histopatologicznego. Zmiany złośliwe opisywano w oparciu o klasyfikację FIGO. W analizie statystycznej guzy graniczne traktowano jak złośliwe. Do oceny istotności statystycznej różnicy stężeń HE4 i CA125, względem złośliwości zmian, użyto testu Mann-Whitney *U*. Dla powyższych testów diagnostycznych, opracowano pole pod krzywą ROC-AUC, które porównano, używając metody Hanley i McNeil. W celu porównania skuteczności stosowanych metod, utworzono dwie grupy obejmujące zmiany złośliwe i niezłośliwe, a także dodatkowo wyodrębniono trzy podgrupy: nowotwory złośliwe w I stopniu zaawansowania wg FIGO, nabłonkowe złośliwe nowotwory jajnika (EOC) oraz niezłośliwe guzy przydatków. Za istotną statystycznie przyjęto wartość *p* poniżej 0,05.

Drugi etap badania, przeprowadzono wykorzystując cztery klasyczne wskaźniki ryzyka złośliwości (RMI 1-4). Analizie poddano 312 pacjentek operowanych od października 2012 do maja 2015 roku, w Katedrze i Klinice Położnictwa, Chorób Kobięcych i Ginekologii

Onkologicznej, II Wydziału Lekarskiego, Warszawskiego Uniwersytetu Medycznego, wśród nich wszystkie 302 pacjentki badane w pierwszym etapie.

Kryteria włączenia do badania obejmowały wiek powyżej 18 lat, ultrasonograficzną ocenę guzów przydatków oraz pomiar CA125 i HE4 nie później niż do pięciu dni od operacji oraz zgodę pacjentki.

Kryteria wykluczenia obejmowały choroby nerek, choroby nowotworowe w wywiadzie, chemio i/lub radioterapię, mięśniaki powyżej 5 cm i brak wyniku badania histopatologicznego zmian przydatków.

Badanie ultrasonograficzne wykonywano aparatem Philips iU22, sondą przezpochwową. Sondę przezbrzuszną stosowano u pacjentek, które nie rozpoczęły współżycia, w przypadku gdy guz „wychodził” poza pole obrazowania endosondy lub w celu wykrycia przerzutów poza obszarem miednicy mniejszej w przypadku podejrzenia zmian złośliwych. W badaniu ultrasonograficznym rejestrowano obecność torbieli wielokomorowych, elementów litych, obustronnych zmian, wodobrzusza oraz przerzutów. Cechy ultrasonograficzne oceniano w oparciu o kryteria Międzynarodowej Grupy Analizy Guzów Jajników IOTA. Przerzuty odległe obejmowały zmiany ogniskowe w nerkach, śledzionie, wątrobie oraz w sieci. W zmianach zlokalizowanych obustronnie, analizie poddawano te o bardziej złożonej echostrukturze. Ultrasonograficzną część wskaźnika RMI (U) wyznaczano na podstawie sumy punktów, przyznawanych za każdą cechę ultrasonograficzną. Jako parametr dodatkowy, do wskaźnika RMI 4 dodano największy wymiar badanej zmiany. Menopauzę definiowano jako brak krwawienia miesięczkowego w ciągu ostatniego roku. Poziom HE4 i CA125 oceniano metodą elektrochemiluminescencji, z tej samej próbki jednym aparatem (Cobas 8000-e602).

Klasyczne wskaźniki RMI 1-3 obejmowały ocenę ultrasonograficzną (U), status menopauzalny (M) oraz poziom CA125 wg wzoru $RMI\ 1-3 = U \times M \times CA125$. Wskaźnik RMI 4 dodatkowo uwzględniał największy wymiar zmiany (S) wg wzoru $RMI\ 4 = U \times M \times CA125 \times S$. Dla wskaźnika RMI 1, wartość U 0,1,3 była przydzielona dla łącznej sumy punktów odpowiednio 0, 1 i ≥ 2 . Dla wskaźników RMI 2 i 4, wartość U 1 i 4, dla łącznej sumy punktów odpowiednio ≤ 1 i ≥ 2 . Dla wskaźnika RMI 3, wartość U 1 i 3, dla łącznej sumy punktów odpowiednio ≤ 1 i ≥ 2 .

Status menopauzalny (M) przyjmował wartość 1 dla kobiet przed menopauzą oraz 3 lub 4 u kobiet po menopauzie w zależności od tego, czy cecha M występowała odpowiednio w modelu RMI 1 i 3 czy RMI 2 i 4. Poziom CA125 stosowano bezpośrednio dla formuły klasycznej RMI 1-4. Wartość S wynosiła 1 lub 2, odpowiednio dla guzów o maksymalnej średnicy <70 mm i ≥ 70 mm. Zmodyfikowaną formułę RMI 1-4 (mRMI 1-4) obliczano w oparciu o klasyczną, stosując w miejsce CA125 wartość HE4 wg wzoru: $mRMI\ 1-3 = U \times M \times HE4$ i $mRMI\ 4 = U \times M \times HE4 \times S$.

Ostateczna ocena guzów przydatków oparta była na pooperacyjnym badaniu histopatologicznym. Stopień zaawansowania zmian złośliwych określano w oparciu o klasyfikację FIGO. W analizie statystycznej guzy graniczne traktowano jak złośliwe. Dla wskaźnika RMI 1-3 punkt odcięcia stanowił 200, dla wskaźnika RMI 4 - 450, zgodnie z wartościami przyjętymi dla tych wskaźników.

Do oceny statystycznej różnicy stężeń HE4 i CA125, względem złośliwości zmian, użyto testu Mann-Whitney *U*. Do analizy wieku pacjentek, statusu menopauzalnego oraz punktacji ultrasonograficznej zmian przydatków stosowano test chi-kwadrat. Dla klasycznych i zmodyfikowanych systemów punktacji opracowano ROC-AUC, w celu porównania przydatności diagnostycznej tych wskaźników w różnicowaniu złośliwych i niezłośliwych zmian przydatków w całej badanej grupie, a także w celu odróżnienia zmian złośliwych w I stopniu zaawansowania wg FIGO od zmian niezłośliwych. Różnice AUC pomiędzy klasycznymi i zmodyfikowanymi wskaźnikami RMI 1-4 porównano, używając metody Hanley i McNeil. Za istotną statystycznie przyjęto wartość *p* poniżej 0,05.

Wyniki

W pierwszym etapie, do badania zakwalifikowano 302 pacjentki. W wieku 18-85 lat, średnio 48,7, z odchyleniem standardowym $SD \pm 16,79$ lat. Przed menopauzą było 188 (62,3%), po menopauzie 114 (37,7%) pacjentek. W ostatecznym badaniu histopatologicznym rozpoznano 50 (16,6%) zmian złośliwych oraz 252 (83,4%) niezłośliwych. Nowotwory złośliwe dotyczyły głównie guzów jajnika ($n=48$ [96%]), oraz dwa przypadki (4%) jajowodu. Stopień klinicznego zaawansowania nowotworu wg FIGO przedstawiał się następująco: IA – 9 (18%) pacjentek, IC – 6 (12%) pacjentek, IIA – 4 (8%) pacjentki, IIC – 1 (2%) pacjentka, IIIA – 1 (2%) pacjentka, IIIB – 2 (4%) pacjentki, IIIC – 24 (48%) pacjentki, IVB – 1 (2%) pacjentka. Jeden przypadek (2%) raka jajowodu rozpoznano w stopniu IC, jak również jeden (2%) w stopniu IIIA zaawansowania klinicznego według FIGO. Większość zmian złośliwych

jajnika, było pochodzenia nabłonkowego (n=46, 95,83%). Pozostałe dwa przypadki dotyczyły ziarniszczaka (2,1%) i mięsaka jajnika (2,1 %).

Testem Mann-Whitney *U* wykazano istotne wyższe stężenie markerów CA125 i HE4 w złośliwych zmianach przydatków, w całej badanej grupie ($p < 0,001$ dla obydwu markerów) przed menopauzą (wartość p odpowiednio 0,008 i $< 0,001$) i po menopauzie ($p < 0,001$ dla obydwu markerów).

ROC-AUC obliczano dla markerów i algorytmu ROMA zarówno dla całej populacji jak i osobno dla grup przed i po menopauzie. Wszystkie testy diagnostyczne znacząco różnicowały zmiany złośliwe i niezłośliwe zarówno dla całej populacji jak i osobno dla grup przed i po menopauzie. Wartość liczbowa AUC dla CA125 wykazująca 95% przedział ufności CI wynosiła 0,838 (0,76-0,917), 0,749 (0,548-0,950) oraz 0,891 (0,816-0,966) odpowiednio dla całej populacji oraz przed i po menopauzie i istotnie statystycznie różnicowała złośliwe i niezłośliwe zmiany przydatków (wartość p odpowiednio $< 0,001$, 0,008 i $< 0,001$). AUC dla HE4 wykazująca 95% CI wynosiła 0,928 (0,885-0,971), 0,867 (0,739-0,996) oraz 0,915 (0,857-0,972) odpowiednio dla całej populacji oraz przed i po menopauzie i istotnie statystycznie różnicowała zmiany złośliwe i niezłośliwe przydatków ($p < 0,001$ dla wszystkich grup). AUC dla algorytmu ROMA wykazująca 95% CI wynosiła 0,928 (0,883-0,972), 0,865 (0,729-1,000) oraz 0,917 (0,857-0,977) odpowiednio dla całej populacji oraz przed i po menopauzie i istotnie statystycznie różnicowała zmiany złośliwe i niezłośliwe przydatków ($p < 0,001$ dla wszystkich grup).

Wartość liczbowa AUC wykazująca 95% CI dla HE4, CA125 i algorytmu ROMA w różnicowaniu zmian złośliwych w I stopniu zaawansowania wg FIGO od niezłośliwych zmian przydatków wynosiła odpowiednio 0,802 (0,695-0,910), 0,559 (0,388-0,731) i 0,789 (0,679-0,898). Marker HE4 i algorytm ROMA znacząco lepiej niż CA125 różnicował guzy złośliwe w I stopniu zaawansowania wg FIGO od niezłośliwych zmian przydatków (wartość $p < 0,001$ dla HE4 i algorytmu ROMA i 0,427 dla CA125).

Wartość liczbowa AUC wykazująca 95% CI dla HE4, CA125 i algorytmu ROMA w różnicowaniu złośliwych nabłonkowych zmian jajnika od niezłośliwych zmian przydatków wynosiła odpowiednio 0,928 (0,882-0,974), 0,858 (0,781-0,935) i 0,929 (0,882-0,976) co oznacza, że wszystkie użyte testy diagnostyczne znacząco różnicowały nabłonkowe zmiany złośliwe jajnika od niezłośliwych zmian przydatków ($p < 0,001$).

Algorytm ROMA i HE4 wykazał lepszą skuteczność niż CA125 w różnicowaniu guzów złośliwych i niezłośliwych w całej badanej grupie (odpowiednio $p=0.043$ i $p=0.043$). Podobnie algorytm ROMA i HE4 były znacząco lepsze niż CA125 w różnicowaniu złośliwych guzów w stopniu I zaawansowania wg FIGO od zmian niezłośliwych w całej badanej grupie (odpowiednio $p=0,025$ i $p=0,017$). Nie wykazano istotnej większej wartości algorytmu ROMA w stosunku do HE4 i CA125 w różnicowaniu EOC i niezłośliwych zmian przydatków (wartość p odpowiednio 0,979 i 0,112).

W drugim etapie, do badania zakwalifikowano 312 pacjentek, z czego 117 (37,5%) było po menopauzie. Wiek pacjentek wyniósł 18-85 lat, średnio $48,5 \pm SD 16,8$. W ostatecznym badaniu histopatologicznym rozpoznano 52 (16,67%) zmian złośliwych. Stopień klinicznego zaawansowania nowotworu wg FIGO przedstawiał się następująco: IA – 9 (17,3%) pacjentek, IC – 7 (13,4%) pacjentek, IIA - 4 (7,6%) pacjentki, IIC – 1 (1,9%) pacjentka, IIIA – 2 (3,8%) pacjentki, IIIB – 2 (3,8%) pacjentki, IIIC - 25 (48%) pacjentek i IVB - 2 (3,8%) pacjentki. Stężenie HE4 i CA125 było znacząco wyższe u pacjentek z guzami złośliwymi ($p<0,001$). Przy użyciu testu chi-kwadrat wykazano znaczącą różnicę w częstości występowania zmian złośliwych u kobiet w starszym wieku ($p<0,001$), po menopauzie ($p<0,001$), przy wyższej ultrasonograficznej punktacji ($p<0,001$) i przy większych średnicach guzów ($p<0,001$).

Wartość liczbowa AUC wykazująca 95% CI dla klasycznych wskaźników RMI 1-4 dla całej grupy wynosiła odpowiednio 0,899 (0,838-0,960), 0,900 (0,840-0,960), 0,895 (0,832-0,957) i 0,908 (0,851-0,966), a dla zmodyfikowanych wskaźników mRMI 1-4 odpowiednio 0,903 (0,844-0,963), 0,929 (0,886-0,972), 0,930 (0,887-0,973) i 0,931 (0,887-0,976).

Wartość liczbowa AUC wykazująca 95% CI w grupie kobiet przed menopauzą dla klasycznych wskaźników RMI 1-4 wynosiła odpowiednio 0,818 (0,654-0,982), 0,798 (0,606-0,99), 0,795 (0,604-0,986) i 0,802 (0,604-1,000), a dla wskaźników mRMI 1-4 odpowiednio 0,839 (0,676-1,000), 0,875 (0,747-1,000), 0,876 (0,748-1,000) i 0,856 (0,707-1,000).

Wartość liczbowa AUC wykazująca 95% CI w grupie kobiet po menopauzie dla klasycznych wskaźników RMI 1-4 wynosiła odpowiednio 0,906 (0,838-0,973), 0,895 (0,822-0,969), 0,896 (0,823-0,970) i 0,906 (0,842-0,971), a dla wskaźników mRMI 1-4 odpowiednio 0,907 (0,841-0,973), 0,923 (0,869-0,978), 0,924 (0,870-0,979) i 0,930 (0,883-0,977).

Zarówno klasyczne jak i zmodyfikowane wskaźniki RMI 1-4 znacząco różnicowały zmiany złośliwe i niezłośliwe dla całej badanej grupy ($p < 0,001$) oraz u kobiet po menopauzie

($p < 0,001$). W grupie pacjentek przed menopauzą, klasyczne wskaźniki RMI 1-4 również znacząco różnicowały zmiany złośliwe ($p < 0,001$ dla RMI 1 i $0,001$ dla RMI 2-4). Podobnie zmodyfikowany model RMI 1-4 znacząco różnicował zmiany złośliwe ($p < 0,001$).

Wartość liczbowa AUC wykazująca 95% CI dla klasycznych wskaźników RMI 1-4 różnicujących guzy złośliwe w I stopniu zaawansowania wg FIGO od zmian niezłośliwych wynosiła odpowiednio 0,686 (0,535-0,837), 0,692 (0,541-0,843), 0,674 (0,519-0,830) i 0,717 (0,570-0,864), a dla wskaźników mRMI 1-4 odpowiednio 0,715 (0,561-0,869), 0,802 (0,690-0,913), 0,803 (0,692-0,914) i 0,801 (0,682-0,920).

Cały klasyczny model RMI 1-4 znacząco różnicował zmiany złośliwe w I stopniu zaawansowania wg FIGO i niezłośliwe zmiany przydatków (wartość p odpowiednio 0,013, 0,010, 0,019 i 0,004). Podobnie, cały zmodyfikowany model RMI 1-4 istotnie różnicował te zmiany (wartość p odpowiednio 0,004, $< 0,001$, $< 0,001$ i $< 0,001$).

AUC dla wszystkich zmodyfikowanych wskaźników mRMI był w całej badanej grupie 312 kobiet wyższy niż AUC dla klasycznych wskaźników RMI 1-4. Niemniej jednak, zgodnie z wartością p według Hanley i McNeil, zarówno każdy klasyczny, jak i odpowiedni zmodyfikowany wskaźnik RMI 1-4, nie wykazał istotnych różnic dla całej badanej grupy (wartość p odpowiednio 0,923, 0,452, 0,368 i 0,540) jak i po uwzględnieniu statusu menopauzalnego- przed menopauzą (wartość p odpowiednio 0,848, 0,471, 0,449 i 0,620) i po menopauzie (wartość p odpowiednio 0,983, 0,543, 0,541 i 0,585), jak również w różnicowaniu zmian złośliwych w I stopniu zaawansowania wg FIGO od zmian niezłośliwych (wartość p odpowiednio 0,784, 0,276, 0,203 i 0,402).

Wnioski

1. Algorytm ROMA i HE4 wykazuje wyższą skuteczność od CA125 w różnicowaniu złośliwych i niezłośliwych zmian przydatków, zarówno w całej grupie jak też w I stopniu zaawansowania wg FIGO.
2. Algorytm ROMA, w porównaniu do HE4 i CA125, nie wykazuje istotnej większej skuteczności w różnicowaniu złośliwych nabłonkowych zmian jajnika i niezłośliwych guzów przydatków.

3. W przedoperacyjnym różnicowaniu wszystkich zmian złośliwych jak też w I stopniu zaawansowania wg FIGO oraz zmian niezłośliwych są przydatne zarówno klasyczne jak i zmodyfikowane wskaźniki RMI 1-4.

4. Zastąpienie CA125 przez HE4 w klasycznym wskaźniku RMI nie zwiększyło istotnie skuteczności w różnicowaniu wszystkich zmian złośliwych oraz zmian złośliwych w I stopniu zaawansowania wg FIGO oraz zmian niezłośliwych przydatków.

SUMMARY

Introduction

Adnexal masses represent a heterogenous group of benign and malignant changes. The choice of proper management depends on medical history, clinical examination, tumor markers and radiological imaging, mainly ultrasound scan and computed tomography. Several mathematical models and scoring systems (combining clinical features, ultrasound parameters and/or tumor markers) have been developed to improve diagnostic performance of the individual parameters.

Objective

The aim of the study was to assess the diagnostic performance of selected mathematical models and scoring systems for presurgical differentiation of adnexal tumors. The following steps were taken:

- 1) comparison of the diagnostic performance of the Risk of Ovarian Malignancy Algorithm (ROMA) with individual CA125 and HE4 tumor markers;
- 2) comparison of the diagnostic performance of the modified Risk of Malignancy Indices (mRMIs) 1-4, which were calculated by replacing CA125 with HE4, with the classical risk of malignancy indices (RMIs) 1-4 for differentiation of malignant and non-malignant adnexal masses.

Material and methods

The prospective study was divided into two stages. A total of 312 patients (302 in the first and additional 10 in the second stage) were analyzed.

In the first stage of the study, ROMA was used in 302 patients with adnexal masses, admitted for surgery to the Clinic of Obstetrics, Gynecology and Gynecological Oncology, II Faculty of Medicine, Medical University of Warsaw, between October 2012 and April 2015.

The following inclusion criteria were used: age older than 18 years, measurement of serum concentration of CA125 and HE4 <5 days before surgical intervention, histopathological results for the adnexal lesion, and patient consent.

The exclusion criteria included pregnancy, renal diseases, history of malignancy, chemotherapy and/or radiotherapy, fibroids of >5 cm in diameter, and lack of histological result for the adnexal tumor.

Serum HE4 and CA125 levels were measured for each patient at the same time and with the same apparatus (Cobas 8000-e602), using an electrochemiluminescence immunoassay. The two logistic regression formulas described by Moore et al., were used to calculate ROMA. The predictive index (PI) was calculated for the pre- and post-menopausal subgroups: $PI = -12.0 + 2.38 \cdot \ln(HE4) + 0.0626 \cdot \ln(CA125)$ for the pre-menopausal subgroup, and $PI = -8.09 + 1.04 \cdot \ln(HE4) + 0.732 \cdot \ln(CA125)$ for the post-menopausal subgroup, with ln standing for natural logarithm in both groups.

The following formula was applied to calculate the risk of malignancy based on ROMA (%): $\% = \exp(PI) / [1/\exp(PI)] \cdot 100$. The cut-off levels were as follows: 35 U/mL for CA125, 70 and 140 pmol/L for HE4, 11.4% and 29.9% for ROMA for high-risk pre- and post-menopausal patients, respectively. The final diagnosis of the adnexal masses was based on the result of postoperative histopathology. FIGO guidelines were used for the staging of the malignant masses. In statistical analysis, borderline tumors were considered as malignant. The Mann–Whitney *U* test was used to assess the statistical difference of mean serum levels of HE4 and CA125 considering malignancy. The receiver operating characteristics area under the curve (ROC-AUC) was constructed for the diagnostic tests and compared using the Hanley&McNeil method. Two groups of patients, with malignant and non-malignant adnexal masses, were formed to compare the diagnostic performance of the tests. Additionally, three subgroups were created: stage I FIGO malignant adnexal masses, epithelial ovarian cancer (EOC), and non-malignant adnexal tumors. The p-value of <0.05 was considered to be statistically significant.

The second stage was conducted to assess the diagnostic performance of the four classical risk of malignancy indices (RMIs 1-4). A total of 312 patients who underwent surgery at the Clinic of Obstetrics, Gynecology and Gynecological Oncology, II Faculty of Medicine,

Medical University of Warsaw, between October 2012 and May 2015, were investigated, including the 302 patients from the first stage of the study.

The inclusion criteria were as follows: age ≥ 18 years, ultrasound assessment of the adnexal mass and CA125 and HE4 taken measurements within 5 days before surgical intervention, and informed consent.

The exclusion criteria included renal disease, history of malignancy, chemotherapy, radiotherapy, presence of fibroids (>5 cm in diameter) and lack of histological assessment of the mass.

The ultrasound examination was performed using Philips iU22 with a transvaginal probe. The transabdominal ultrasound was performed in virgin patients, or when the mass could not be visualized by a transvaginal ultrasound, or to detect metastases outside the pelvis in cases with suspected malignancy. The following ultrasound findings were considered: multilocular cyst, solid areas, bilateral lesions, ascites, and metastases. The definitions of multilocular cyst, solid areas, and ascites were consistent with the terms and definitions established by the International Ovarian Tumor Analysis (IOTA) group. Distant metastases were defined as focal lesions in renal, splenic, and hepatic parenchyma or in the omentum. If bilateral lesions were noted, only data regarding the lesion with a more complex echostructure were included in the statistical analysis. One point was assigned for each ultrasound finding and the sum was used to determine the value for the RMI ultrasound score (U). The maximum diameter of the lesion (S) was considered as an additional ultrasound parameter in RMI 4. Menopause was defined as the absence of menstruation for at least 1 year. Serum CA125 and HE4 levels were measured at the same time using electrochemiluminescence immunoassay (with Cobas 8000-e602).

The formulas of classical RMIs 1–3 used the ultrasound score (U), menopausal status (M), and CA125 serum levels (RMIs 1–3 = $U \times M \times CA125$), while classical RMI 4 also included the maximum diameter of the lesion (S) (RMI 4 = $U \times M \times CA125 \times S$). For RMI 1, a U-value of 0, 1, and 3 was assigned when total ultrasound scores were 0, 1, and ≥ 2 points, respectively. For RMIs 2 and 4, U-values of 1 and 4 were assigned when the total ultrasound scores were ≤ 1 and ≥ 2 points, respectively. For RMI 3, U-values of 1 and 3 were assigned when total ultrasound scores were ≤ 1 and ≥ 2 points, respectively.

The menopausal (M) status had a value of 1 in pre-menopausal patients and a value of 3 or 4 in post-menopausal patients, depending on whether M was included in RMIs 1 and 3 or RMIs 2 and 4, respectively. CA125 serum levels were used directly in the formulas of RMIs 1–

4. S-values were 1 and 2 for maximum lesion diameters of <70 mm and ≥ 70 mm, respectively. The corresponding modified RMIs 1–4 (mRMIs 1–4) were calculated using the same formulas after replacing CA125 serum levels with HE4 as follows: mRMIs 1–3 = $U \times M \times HE4$ and mRMI 4 = $U \times M \times HE4 \times S$.

The final diagnosis of the adnexal masses was based on postoperative histopathology. The staging of the malignant masses was performed using the FIGO guidelines. In the statistical analysis, borderline tumors were considered as malignant. A cut-off of 200 and 450 was set for RMIs 1–3 and RMI 4, respectively, in accordance with the original values of each index.

The Mann–Whitney U test was used to assess the differences in CA125 and HE4 distribution, whereas the chi-square test was used to assess the differences in the distribution of age, menopausal status, and ultrasound score versus the malignancy status of the adnexal mass. The area under the ROC curve (AUC) was constructed for both, classical RMIs and mRMIs 1–4 to differentiate malignant from non-malignant adnexal masses in the whole sample. These measures were also used for the differentiation of malignant stage I (FIGO) adnexal tumors from non-malignant adnexal masses. The Hanley&McNeil method was used to assess the difference between the AUCs of classical RMIs and mRMIs. The p-value of <0.05 was considered to be statistically significant.

Results

A total of 302 patients were included in the first stage of the study. Patient age was 18–85 years (mean: 48.7 years, SD: 16.79 years). Pre-menopausal patients comprised the majority of the study population ($n=188$ [62.3%]) and only 114 (37.7%) patients were post-menopausal. The final histopathological examination revealed 252 (83.4%) cases of non-malignant and 50 (16.6%) cases of malignant adnexal pathologies. The vast majority of the malignant tumors ($n = 48$ [96%]) were of ovarian origin; two (4%) were fallopian tube malignancies. The number of patients with each FIGO stage of malignant ovarian pathology was as follows: stage IA - 9 (18%); IC - 6 (12%); IIA - 4 (8%); IIC - 1 (2%); IIIA - 1 (2%); IIIB - 2 (4%); IIIC - 24 (48%); and IVB - 1 (2%). Tubal malignancies included one case of stage IC (2%) and one of IIIA (2%). Most malignant ovarian cases were of epithelial origin ($n = 46$; 95.83%). There was also one case (2.1%) of ovarian folliculoma and one case (2.1%) of ovarian sarcoma.

The Mann–Whitney U test showed that both, CA125 and HE4 showed significantly higher serum levels in patients with malignant adnexal masses in the entire study population

(p-value of <0.001 for both markers) among the pre-menopausal (p-values of 0.008 and <0.001, respectively) as well as post-menopausal subjects (p-value of <0.001 for both markers).

A ROC-AUC was computed both, for tumor markers and for ROMA for the entire study group, as well as for the pre- and post-menopausal subgroups. All diagnostic tests significantly differentiated malignant adnexal tumors from non-malignant adnexal tumors in the study group, as well as the pre- and post-menopausal subgroups. The numerical values of AUC for CA125 (with 95% confidence interval; (CI)) were 0.838 (0.76–0.917), 0.749 (0.548–0.950) and 0.891 (0.816–0.966) in the entire study group, pre- and post-menopausal subgroups, respectively, and significantly differentiated malignant from non-malignant adnexal masses (p-values of <0.001, 0.008 and <0.001, respectively). The numerical values of AUC for HE4 (with 95% CI) were 0.928 (0.885–0.971), 0.867 (0.739–0.996) and 0.915 (0.857–0.972) in the entire study group, pre- and post-menopausal subgroups, respectively, and significantly differentiated malignant from non-malignant adnexal masses (p-value of <0.001 in each group). The numerical value of AUC for ROMA (with 95%CI) were 0.928 (0.883–0.972), 0.865 (0.729–1.000) and 0.917 (0.857–0.977) in the entire study group, pre- and post-menopausal subgroups, respectively, and significantly differentiated malignant from non-malignant adnexal masses (p-value of <0.001 in each group).

The numerical values of AUC (with 95% CI) for HE4, CA125 and ROMA for differentiation of malignant stage I FIGO from non-malignant adnexal tumors were 0.802 (0.695–0.910), 0.559 (0.388–0.731) and 0.789 (0.679–0.898), respectively. Both, HE4 and ROMA were significantly better than CA125 at differentiating stage I FIGO malignant tumors from non-malignant adnexal tumors (p-values of <0.001 for HE4 and ROMA and 0.427 for CA125).

The numerical values of AUC (with 95% CI) for HE4, CA125 and ROMA for differentiation of epithelial ovarian cancer from non-malignant adnexal tumors were 0.928 (0.882–0.974), 0.858 (0.781–0.935) and 0.929 (0.882–0.976), respectively, and all these diagnostic tests significantly differentiated epithelial ovarian cancer from non-malignant adnexal tumors (p-value <0.001 for all tests).

ROMA and HE4 were significantly more effective in the differentiation of malignant from non-malignant tumors in the entire study group (p=0.043 and p=0.043, respectively) as compared to CA125. Similarly, ROMA and HE4 were significantly better for the differentiation of stage I FIGO malignant from non-malignant adnexal masses (p-values of 0.025 and 0.017,

respectively) as compared to CA125. However, ROMA was not significantly better as compared to HE4 or CA125 for the differentiation of EOC from non-malignant adnexal masses (p-values of 0.979 and 0.112, respectively).

A total of 312 women were included in the second stage of the study, and 117 (37.50%) patients were post-menopausal at the time of the assessment. Patient age ranged from 18 to 85 years (mean: 48.5 years; SD: 16.8 years). Malignant pathologies were diagnosed in 52 women (16.67%). The number of patients with each stage of malignant adnexal pathology (FIGO) was as follows: stage IA - 9 (17.3%); IC - 7 (13.4%); IIA - 4 (7.6%); IIC - 1 (1.9%); IIIA - 2 (3.8%); IIIB - 2 (3.8%); IIIC - 25 (48%); and IVB - 2 (3.8%). HE4 and CA125 serum levels were significantly higher among patients with malignant adnexal masses (p-value of <0.001). The chi square test showed statistical significant differences for the occurrence of malignancy in terms of advanced age (p-value of <0.001), among post-menopausal patients (p-value of <0.001), higher ultrasound score (p-value of <0.001) and larger maximum tumor diameter (p-value of <0.001).

The numerical values of AUC (with 95% CI) of classical RMIs 1-4 in the entire study population were 0.899 (0.838–0.960), 0.900 (0.840–0.960), 0.895 (0.832–0.957) and 0.908 (0.851–0.966), respectively. The corresponding mRMIs 1-4 were 0.903 (0.844–0.963), 0.929 (0.886–0.972), 0.930 (0.887–0.973) and 0.931 (0.887–0.976), respectively.

In the pre-menopausal group, the numerical values of AUC (with 95% CI) of classical RMIs 1-4 were 0.818 (0.654–0.982), 0.798 (0.606–0.99), 0.795 (0.604–0.986) and 0.802 (0.604–1.000), respectively. The corresponding mRMIs 1-4 values were 0.839 (0.676–1.000), 0.875 (0.747–1.000), 0.876 (0.748–1.000) and 0.856 (0.707–1.000), respectively.

In the post-menopausal group, the numerical values of AUC (with 95% CI) of classical RMIs 1-4 were 0.906 (0.838–0.973), 0.895 (0.822–0.969), 0.896 (0.823–0.970) and 0.906 (0.842–0.971), respectively. The corresponding mRMIs 1-4 were 0.907 (0.841–0.973), 0.923 (0.869–0.978), 0.924 (0.870–0.979) and 0.930 (0.883–0.977), respectively.

All classical RMIs and mRMIs significantly differentiated malignant from non-malignant adnexal tumors in the entire study population (p-value of <0.001), and among post-menopausal patients (p-value of <0.001). In premenopausal group, all classical RMIs 1-4 also significantly differentiated malignancies (p-value of <0.001 for RMI 1 and 0.001 for RMIs 2-4). All modified RMIs 1-4 also had significant statistical differences (p-value of <0.001).

The numerical values of AUC (with 95% CI) for classical RMIs 1-4 for differentiation of malignant stage I (FIGO) tumors from non-malignant adnexal tumors were: 0.686 (0.535–0.837), 0.692 (0.541–0.843), 0.674 (0.519–0.830) and 0.717 (0.570–0.864), respectively. As for mRMIs 1-4, the values were 0.715 (0.561–0.869), 0.802 (0.690–0.913), 0.803 (0.692–0.914) and 0.801 (0.682–0.920), respectively.

All classical RMIs 1-4 significantly differentiated stage I (FIGO) malignant from non-malignant adnexal tumors (p-values of 0.013, 0.010, 0.019 and 0.004, respectively). Modified RMIs 1-4 also significantly differentiated malignant tumors stage I (FIGO) from non-malignant adnexal tumors (p-values of 0.004, <0.001, <0.001 and <0.001, respectively).

The AUCs for all mRMIs were higher than those of classical RMIs 1-4 in the entire study sample of 312 women. Nonetheless, taking into consideration the Hanley&McNeil *p* values, AUCs of the corresponding classical RMIs and mRMIs 1–4 were not significantly different in the entire study sample (p-values of 0.923, 0.452, 0.368 and 0.540, respectively) or in either of the subgroups, i.e. pre-menopausal (p-values of 0.848, 0.471, 0.449 and 0.620, respectively) or post-menopausal (p-values of 0.983, 0.543, 0.541 and 0.585, respectively) patients, as well as in the differentiation of malignant stage I (FIGO) from non-malignant adnexal masses (p-values of 0.784, 0.276, 0.203 and 0.402, respectively).

Conclusions

1. In the whole group, ROMA and HE4 were more useful than CA125 for the differentiation of malignant from non-malignant adnexal tumors, which also proved to be true for the differentiation of stage I (FIGO) malignant from non-malignant adnexal tumors.
2. ROMA is not superior to tumor markers HE4 and CA125 for the differentiation of EOC from non-malignant adnexal tumors.
3. Classical RMIs and mRMIs 1–4 are useful for pre-surgical differentiation of all malignant and stage I (FIGO) malignant from non-malignant adnexal tumors.
4. Replacing CA125 with HE4 in classic RMIs 1-4 formulas did not improve the diagnostic performance of these indices for the differentiation of all malignant and stage I (FIGO) malignant tumors from non-malignant adnexal tumors.