



Wrocław, 22 września 2024

RECENZJA

ROZPRAWY DOKTORSKIEJ MGR KATARZYNY KWIATKOWSKIEJ

ZATYTUŁOWANEJ „OCENA PRZYDATNOŚCI OZNACZENIA WYBRANYCH MARKERÓW WYDZIELANYCH PRZEZ KOMÓRKI ŚRÓDBŁONKA NACZYNIOWEGO ORAZ TKANKĘ TŁUSZCZOWĄ JAKO CZYNNIKÓW PROGNOSTYCZNYCH W GRUPIE PACJENTEK Z PIERWOTNYM RAKIEM PIERSI” WYKONANEJ W KATEDRZE PATOFIZJOLOGII COLLEGIUM MEDICUM IM. LUDWIKA RYDYGIERA W BYDGOSZCZY POD KIERUNKIEM PROMOTOR DR HAB. BARBARY CIASTEK

1. Podstawa formalno-prawna opracowania recenzji

Przedmiotem recenzji jest dysertacja doktorska pani mgr Katarzyny Kwiatkowskiej zatytułowana „Ocena przydatności oznaczenia wybranych markerów wydzielanych przez komórki śródbłonka naczyniowego oraz tkankę tłuszczową jako czynników prognostycznych w grupie pacjentek z pierwotnym rakiem piersi” wykonana pod kierunkiem dr hab. Barbary Ciastek. Podstawą formalną opracowania recenzji jest pismo z dnia 11 lipca 2024 (nr LADL.5201.1655.2024) od Przewodniczącego Rady Dyscypliny Nauki Medyczne prof. dr hab. Dariusza Grzanki, w sprawie powierzenia mi oceny rozprawy. Recenzja ma na celu ustalenie, czy rozprawa spełnia wymogi określone w art. 187 Ustawy z dnia 20 lipca 2018 roku *Prawo o szkolnictwie wyższym i nauce* (Dz. U. z 2021 poz. 478 ze zm.).

2. Formalna ocena pracy

Przedstawiona do oceny praca ma formę przewidzianą w ustawie z dnia 20 lipca 2018 roku *Prawo o szkolnictwie wyższym i nauce* art. 187 ust. 3., czyli stanowi (poprzedzony wstępem) „zbiór opublikowanych i powiązanych tematycznie artykułów naukowych”. W skład przedstawionego do oceny zbioru prac wchodzi trzy publikacje oryginalne. Do każdej z prac dołączono zestaw oświadczeń autorów publikacji, w których każdy z nich przedstawił swój wkład w powstanie pracy. Z oświadczeń wynika, że udział Doktorantki w przygotowaniu tych publikacji był duży i wynosił 40% dla publikacji 1, oraz po 75% dla publikacji nr 2 i 3.

3. Ocena merytoryczna wstępu

Wstęp do rozprawy jest obszerny i liczy 48 stron. Zawiera on podstawowe informacje na temat patogenety, epidemiologii diagnostyki oraz leczenia nowotworów piersi. W dalszych częściach wstępu Autorka opisała budowę i funkcje tkanki tłuszczowej, śródbłonka naczyniowego oraz proces hemostazy

Prodziekan
Wydziału Lekarskiego
ds. Nauki

prof. dr hab. Tomasz Grzybowski

Należy podkreślić, że czytanie wstępu jest ułatwione bardzo starannie przygotowanymi ilustracjami. Sposób przygotowania wstępu sugeruje, że treścią rozprawy będą biologiczne aspekty interakcji między guzami piersi, a śródbłonkiem naczyniowym, układem krzepnięcia i tkanką tłuszczową.

Wstęp został przygotowany dość starannie, należy jednak zwrócić uwagę na kilka istotnych błędów. W ostatnim zdaniu na stronie 11 Autorka podaje, że jedna na osiem obecnie żyjących kobiet rozwinie w ciągu swojego życia ten typ nowotworu (nowotwór piersi). Autorka powołuje się w tym zdaniu na publikację nr 10 ze spisu piśmiennictwa, opublikowaną w 2022 roku, zatytułowaną *Breast Cancer Statistics*. Następnie na stronie 12, Autorka pisze, że „zachorowania na raka piersi sięgają aż 99% kobiet”, co stoi w sprzeczności ze zdaniem z poprzedniej strony. W tym miejscu Autorka powołuje się na publikację nr 14 ze spisu piśmiennictwa, zatytułowaną *Breast Cancer-Epidemiology, Classification, Pathogenesis and Treatment (Review of Literature)*. W publikacji tej w rozdziale *Sex* napisano „*The vast majority of cases of breast cancer, reaching 99%, occur in women*”. Znaczenie tego zdania w cytowanej pracy jest odmienne od sposobu w jaki zinterpretowała je Autorka rozprawy doktorskiej.

Na stronie 28 w opisie leku Trastuzumab (Herceptin), Autorka błędnie napisała, że lek ten (przeciwciało) wiąże się z domeną kinazową receptora (dokładnie napisano „domeną receptora kinazy tyrozynowej”). Autorka powołuje się w tym miejscu na publikację 65, zatytułowaną *Chemotherapeutic agents for the treatment of metastatic breast cancer: An update*. Tymczasem w pracy tej, w opisie leku napisano „herceptyna wiąże się z domeną przybłonową receptora kinazy tyrozynowej występującą w HER 2” („*Herceptin binds to the juxtamembrane domain of the tyrosine kinase receptor found in HER 2*”). Domena przybłonowa jest na zewnątrz komórki, przez co jest dostępna dla przeciwciała. Natomiast domena kinazowa jest wewnątrz komórki, przez co przeciwciało nie może się z nią związać.

W tabeli 1 na stronie 19 niektóre dane zostały oznaczone gwiazdkami, nie wyjaśniono nigdzie co te gwiazdki oznaczają. Ponadto wstęp nie jest wolny od błędów gramatycznych, których szczególne nagromadzenie jest na stronie 36.

4. Ocena merytoryczna publikacji wchodzących w skład rozprawy

Na stronie 58 wstępu przedstawiono cel rozprawy. Była nim „ocena stężeń markerów produkowanych przez komórki śródbłonka naczyniowego ... oraz tkankę tłuszczową ... w grupie pacjentek z pierwotnym, jednostronnym rakiem piersi bez przerzutów odległych”. Wydaje się, że w tym zdaniu powinno się raczej użyć sformułowania „pomiar stężeń”, a nie „ocena stężeń”. Dalej napisano, że „kolejnym celem pracy było określenie związków między tymi parametrami a rodzajem zastosowanego leczenia w przypadku markerów porównywanych przed i po leczeniu. Ponadto, celem pracy była również ocena przydatności tych parametrów w przewidywaniu występowania choroby i jej nawrotów...”.

Na stronach 60-61 Autorka przedstawiła swój wkład w powstanie trzech publikacji, które składają się na rozprawę. Z jej deklaracji wynika, że jej udział w pracy polegał w każdym przypadku na przeprowadzeniu analizy statystycznej, interpretacji wyników tej analizy, a następnie na zredagowaniu tekstów publikacji. Z deklaracji tej wynika, że w istocie jest to rozprawa doktorska ze statystyki, a nie jak sugerował wstęp do rozprawy, z biologii nowotworów. Z oświadczeń pozostałych Autorów publikacji wynika, że klasyfikacją i leczeniem pacjentek oraz wykonaniem pomiarów zajmowały się inne osoby. Osoby te swój udział w powstaniu prac oceniły jako bardzo skromny, więc zdecydowana większość odpowiedzialności za treść prac spada na Autorkę rozprawy doktorskiej.

Publikacja 1

Publikacja 1 została ogłoszona w 2020 roku w czasopiśmie *Neoplasma*. W tej pracy Autorka rozprawy doktorskiej jest drugim autorem, podczas gdy pierwszym autorem jest Pani Promotor rozprawy doktorskiej. Jak wynika z oświadczenia, udział Pani Promotor w powstaniu publikacji polegał na sformułowaniu koncepcji badania i hipotezy badawczej oraz na końcowej akceptacji manuskryptu. Tymczasem Autorka rozprawy doktorskiej przeprowadziła przegląd piśmiennictwa, analizę statystyczną, zinterpretowała wyniki, sformułowała wnioski z badań oraz brała udział w redagowaniu manuskryptu. Wydaje się więc, że za sposób przeprowadzenia analizy statystycznej odpowiada Autorka rozprawy doktorskiej.

Badanie opisane w publikacji polegało na starannym wyselekcjonowaniu pacjentek z nowotworami piersi na wczesnym etapie choroby, równie starannym wybraniu grupy kontrolnej, wykonaniu pełnej diagnostyki związanej z oceną zaawansowania nowotworów, oraz dodatkowo na wykonaniu pomiarów BMI, oraz oznaczeniu stężeń tkankowego aktywatora plazminogenu (t-PA) i inhibitora aktywacji plazminogenu typu 1 (PAI-1). Następnie pacjentki były leczone zgodnie z aktualnymi wytycznymi. Po operacji wynik leczenia obserwowano przez 28 do 40 miesięcy. Zgodnie z oświadczeniami za całą tę część badania odpowiadali dr n. med. Piotr Rhone oraz mgr Sylwia Bielawska.

Już w streszczeniu publikacji napisano, że celem pracy była ocena przydatności czterech parametrów do przewidywania dalszego postępu choroby. Te parametry to stężenia w surowicy krwi obwodowej t-PA i PAI-1 oraz dwa współczynniki t-PA/PAI-1 oraz PAI-1/t-PA. Powód dla którego analizuje się oba te współczynniki, zamiast wybrać jeden z nich, jest zagadkowy, ponieważ $(t-PA/PAI-1)$ równe jest $(PAI-1/t-PA)^{-1}$. Jeżeli jeden z nich byłby przydatny prognostycznie, to drugi siłą rzeczy też musiałby taki być.

Początkowo porównywane były cztery wspomniane powyżej parametry między grupami chorych i zdrowych. Z tabeli 3 wynika, że jeden z parametrów, stężenie PAI-1 w surowicy, był istotnie wyższy u pacjentek niż u zdrowych, przy czym wartość prawdopodobieństwa dotyczącego hipotezy zerowej dla tego porównania (p) równa jest 0,0494. W tabeli 3 pojawia się wartość 33,91 ng/ml PAI-1. Jest to wartość

średkowa (*median*) w grupie kobiet zdrowych. Ta sama wartość wymieniona jest w publikacji jeszcze kilka razy. W streszczeniu zamieszczono następujące zdanie: „stężenie PAI-1 ma najdokładniejszą wartość prognostyczną z punktem odcięcia na poziomie 33,91 ng/ml, z 90% czułością i 36% swoistością, co pozwala na rozróżnienie pomiędzy grupą kontrolną i chorych na raka” („*the PAI-1 concentration demonstrates the most accurate prognostic value with the cut-off point at 33.91 ng/ml, with 90% sensitivity and 36% specificity, which discriminates between controls and cancer patients*”). Na stronie 1150 napisano: “Korzystając z wartości odcięcia Indeksu Youdena, zidentyfikowaliśmy stężenie PAI-1 w osoczu wynoszące 33,91 ng/ml o czułości 90% i swoistości 36% jako najlepszą graniczną wartość, według której można rozróżnić zdrowe osoby od chorych na raka piersi (rys. 1)” („*Using the Youden Index cut-off value, we identified a plasma PAI-1 concentration of 33.91 ng/ml with a sensitivity of 90% and a specificity of 36% as the best cut-off value to discriminate between healthy individuals from breast cancer cases (Figure 1)*”). W dyskusji ta wartość pojawia się w takim samym kontekście, jak w streszczeniu publikacji. Jest oczywiste, że wartość średkowa zmierzona wśród osób zdrowych, nie powinna być jednocześnie wartością graniczną rozróżniającą osoby zdrowe od osób chorych, ponieważ połowa zdrowych osób będzie zakwalifikowana jako chore. Swoistość takiego testu będzie wynosić 50%, nie zaś jak błędnie stwierdzono w artykule 36%. Nie potrafię stwierdzić, gdzie został popełniony błąd, jednak błąd ten jest powielany zarówno w publikacji, jak i we wnioskach z rozprawy doktorskiej.

Następnie przeanalizowano rozkład badanych parametrów w grupie pacjentek w zależności od wieku (<55 lat oraz ≥55lat), a wyniki przedstawiono w tabeli 4. Okazało się, że pacjentki starsze mają istotnie wyższy poziom PAI-1 w surowicy, niż młodsze. To odkrycie ma znaczenie jedynie informacyjne, gdyż wiek pacjentek da się łatwo określić bez badania krwi. Przedstawiono też rozkład badanych parametrów w zależności od BMI pacjentek, mimo że rozkład ten okazał się równomierny.

W kolejnym etapie te same parametry (stężenia w surowicy krwi obwodowej t-PA i PAI-1 oraz dwa współczynniki t-PA/PAI-1 oraz PAI-1/t-PA) analizowano już tylko u chorych, w zależności od rozmaitych cech nowotworów (tabela 6). Porównywano je w zależności od (1) rozmiaru guza, (2) stopnia zaawansowania choroby wg WHO, (3) strony ciała w której znajdował się guz, (4) stopnia zaawansowania wg Elston-Ellis; (5) zajęcia węzłów chłonnych przez guz, (6) typu histologicznego, (7) typu molekularnego oraz (8) poziomu występowania antygenu Ki-67. Mamy tu 8 testów wykonanych na tym samym dokładnie zbiorze danych. W takiej sytuacji występuje w analizie statystycznej problem porównań wielokrotnych, wobec czego należy zastosować jakieś narzędzie przeciwdziałające temu problemowi. Często stosowana jest w takich przypadkach poprawka Bonferroniego, polegająca na zmniejszeniu przyjętego *a priori* poziomu istotności każdego z powiązanych testów wprost proporcjonalnie do ich liczby. W tym przypadku zamiast $p < 0,05$, za istotny można uznać dopiero poziom $p < 0,00625$. Taki poziom istotności nie został osiągnięty w żadnym z porównań. Tymczasem Autorzy twierdzą, że „wyższy potencjał fibrynolityczny (t-PA/PAI-1, *przyp. rec.*) zaobserwowano w nowotworach w stopniu T2 i IIA+IIB, a wyższy

profil antyfibrynolityczny (PAI-1/t-PA, *przyp. rec.*) w guzach w stopniu T1 i IA+IB (odpowiednio $p=0,0171$, $p=0,0189$)” („*higher fibrinolytic potential in T2 and IA+IB cancers was observed but a higher anti-fibrinolytic profile in T1 and IA+IB tumors was noted ($p=0.0171$, $p=0.0189$, respectively)*”). Jak można się spodziewać, skoro jeden z tych współczynników jest zdefiniowany jako odwrotność drugiego, gdy jeden z nich jest istotnie wyższy w danej grupie, to ten drugi jest istotnie niższy, przy czym p oczywiście jest identyczne. Niestety, Autorzy nie zauważyli tej prawidłowości.

Warto zastanowić się nad określeniami, których Autorzy artykułu użyli dla opisu współczynników t-PA/PAI-1 oraz PAI-1/t-PA. Współczynnik t-PA/PAI-1 został nazwany „potencjałem fibrynolitycznym”, podczas gdy PAI-1/t-PA „profilem antyfibrynolitycznym”. Zgodnie z przyjętą w literaturze naukowej nomenklaturą symbolem „t-PA/PAI-1” określa się kompleksy t-PA/PAI-1 (*t-PA/PAI-1 complex*, np. <https://doi.org/10.1161/01.STR.31.1.26>). W zapisie tym „/” nie jest zatem symbolem dzielenia. Tworzenie przez białka t-PA oraz PAI-1 kompleksów zmniejsza pulę wolnego i aktywnego t-PA w organizmie. Ponieważ t-PA jest enzymem przekształcającym plazminogen do plazminy i w ten sposób aktywującym fibrynolizę, można by się zgodzić, że stężenie wolnego t-PA świadczy o „potencjale fibrynolitycznym”, a stężenie kompleksów t-PA/PAI-1 o potencjale „antyfibrynolitycznym” badanej osoby. Natomiast podzielone przez siebie wartości stężeń wolnych t-PA oraz PAI-1, tak czy na wspak, nie świadczą o żadnym z tych potencjałów.

Następnie w pracy opisano jej najważniejszą część, czyli ocenę wartości prognostycznej badanych czterech parametrów. Do tego celu posłużono się krzywymi charakterystyki operacyjnej odbiornika (*receiver operating characteristic curves*; ROC). Krzywe ROC są stosowane w celu określania jakości testu przewidującego przynależność obiektu do jednej z dwóch klas, w zależności od kolejnych wartości progowych (*thresholds*). Krzywe wykreśla się w kwadracie o bokach 1 w sposób następujący: dla każdej wartości progowej na wykresie umieszcza się punkt o współrzędnych (x,y) gdzie y jest frakcją wyników prawdziwie pozytywnych (czyli czułością testu), zaś x jest frakcją wyników fałszywie pozytywnych (1 minus swoistość) przy tej wartości progowej. Następnie oblicza się pole pod wykreśloną krzywą (*area under the curve*; AUC). Dla testu idealnego AUC równa się 1 , a dla testu, w którym zamiast przeprowadzenia pomiarów, wynik się losuje AUC wynosi $0,5$ (wykres dla testu losowego jest przekątną kwadratu między punktami $0,0$ i $1,1$). Każdy test, dla którego $AUC < 0,5$ jest gorszy niż losowe przydzielanie do jednej z dwóch kategorii.

W **publikacji 1** zamieszczono dwie krzywe ROC, na rysunkach 1 i 2. Z opisu w tekście na stronach 1150 i 1154 wynika, że rysunek 1 przedstawia wykres przydatności badania czterech badanych parametrów do oceny śmiertelności choroby, tymczasem w legendzie rysunku napisano, że ocenie podlegała użyteczność testów do odróżniania zdrowych od chorych. Nie wiadomo, które z tych stwierdzeń jest prawdziwe. Na rysunku 2 przedstawiono krzywe ROC dla oceny przydatności badanych parametrów w przewidywaniu nawrotów choroby.

Autorzy pokazują, że o ile wyznaczanie parametru t-PA/PAI-1 dla oceny śmiertelności choroby (lub odróżniania zdrowych od chorych) jest nieco lepsze od losowania, to wyznaczanie PAI-1/t-PA daje już wyniki gorsze niż losowanie. Z jakiegoś powodu dla oceny nawrotów choroby jest odwrotnie. Z konstrukcji ROC dla współczynników t-PA/PAI-1 i PAI-1/t-PA oraz przyjętego przez Autorów założenia, że wartości wyższe niż progowa oznaczają złe rokowanie, a wartości niższe niż progowa dobre rokowanie, wynika że krzywe muszą być symetryczne względem środka kwadratu, a suma ich AUC będzie zawsze równa 1. Ponieważ jak wykazano powyżej, odpowiednio zastosowane oba współczynniki powinny mieć identyczną moc predykcyjną, wyniki uzyskane przez Autorów są rezultatem nieprzemyślanego wprowadzenia danych do programu statystycznego. O ile można się pomylić podczas wprowadzania danych do tabeli, to po uzyskaniu absurdalnego wyniku, powinno się przeprowadzić refleksję nad tym, do czego ma służyć proponowany test. Niestety nic takiego nie nastąpiło w przypadku tej pracy.

Kolejne rysunki przedstawiają estymatory Kaplana-Meiera czasu wolnego od progresji choroby, dla progów 5 ng/ml t-PA (rysunek 3), 30 ng/ml PAI-1 (rysunek 4) oraz 7.5 współczynnika PAI-1/t-PA (rysunek 5). Na wykresach tych osie odciętych opisano jako czas od randomizacji (*Time of randomization*), czyli losowego przydzielania pacjentek do grup mimo, że w opisanych badaniach nie występowała w ogóle randomizacja. Prawdopodobnie powinno być tam napisane „czas od diagnozy”, lub „czas od leczenia” w zależności od tego jaki czas faktycznie był brany pod uwagę. Wykresy na rysunkach 3 oraz 5 są identyczne, jest to możliwe przy odpowiednio dobranych progach.

W dyskusji pracy Autorzy stwierdzają, że „nasze badanie wskazuje, że bazowe stężenia t-PA i PAI-1, mogą być stosowane jako właściwe, łatwe do zastosowania wskaźniki prognostyczne obok takich parametrów nowotworu, jak średnica guza, stopień zaawansowania TNM, lub status węzłów chłonnych”. („*our study indicates that the baseline concentration of t-PA and PAI-1 may be used as appropriate, easy-applicable prognostic indicators next to tumor-related determinants such as tumor diameter, TNM staging system, or nodal status*”). Przedstawione wyniki nie upoważniają do takiego stwierdzenia.

Publikacja 2

Publikacja 2 ukazała się w roku 2022 w czasopiśmie *Life*. Na stronie 78 rozprawy, we wprowadzeniu do tej publikacji, napisano że jej celem było zbadanie wartości prognostycznej stężeń adiponektyny oraz leptyny w surowicy krwi przed i po leczeniu pacjentek z nowotworami piersi. W badaniu wzięło udział 70 pacjentek ze zdiagnozowanym luminalnym rakiem piersi, na wczesnym etapie rozwoju. Wszystkie pacjentki leczono chirurgicznie, a następnie radioterapią, brachyterapią, chemioterapią, terapią hormonalną lub przeciwciałami. Przed terapią oraz po zakończonej terapii oznaczano w surowicy krwi pacjentek stężenia leptyny i adiponektyny. Następnie (przez okres do 6 lat) obserwowano, czy pacjentki pozostawały zdrowe, czy też wystąpiły nawroty choroby nowotworowej, w tym metastazy lub śmierć.

W badaniu analizowano stężenia leptyny u pacjentek podzielonych na grupy pod względem 16 parametrów (tabela 2). Mamy tutaj znowu do czynienia z problemem porównań wielokrotnych, co oznacza, że w tym przypadku zamiast $p < 0,05$, za istotny można uznać dopiero poziom $p < 0,003125$, czego Autorzy nie wzięli pod uwagę. Taki poziom istotności został osiągnięty dla niektórych z przedstawionych porównań. Z przedstawionych danych wynika, że stężenie leptyny w surowicy krwi pacjentek było z reguły wyższe po zakończeniu terapii, niż przed jej rozpoczęciem. Wyższe było też u pacjentek z $BMI \geq 25$ niż u tych z $BMI < 25$. To badanie nie ma jednak znaczenia prognostycznego, ponieważ zarówno BMI, jak i fakt przebycia terapii można ustalić w prosty sposób, bez pomiaru stężenia leptyny.

W tabelach 3 oraz 4 przedstawiono stężenia leptyny w grupach pacjentek podzielonych w zależności od terapii, którą przebyły. Rozkład stężeń leptyny w tych grupach był równomierny. W tabeli 5 przedstawiono rozkłady stężeń adiponektyny, z podziałem analogicznym do tabeli 2 oraz z analogicznym błędem metodologicznym związanym z porównaniami wielokrotnymi. Ta analiza wykazała, że najczęściej stężenia adiponektyny są wyższe przed terapią, niż po. Zastanawiające jest, że o ile tabele ze stężeniami leptyny przedstawiają wartości środkowe (*median*), to tabela ze stężeniami adiponektyny przedstawia wartości średnie (*mean*). W opisie wyników, ani w opisie metod statystycznych nie został wyjaśniony taki wybór prezentacji danych. Na rysunkach 2 oraz 3 przedstawiono tzw. heatmapy (*heatmaps*) korelacji między stężeniami leptyny i adiponektyny, a parametrami według których dzielono pacjentki na grupy wraz z współczynnikami korelacji rang Spearmana. Są to podobne wyniki do tych pokazanych w tabelach 2 i 5.

W rozdziale 3.5 zaczyna się najistotniejsza część pracy, czyli analiza prognostycznej przydatności badania stężeń leptyny i adiponektyny. W pierwszym zdaniu napisano, że na rysunku 4 krzywe ROC zostały wykreślone w celu oceny przydatności badania stężeń adipokin (leptyny i adiponektyny) przed terapią do przewidywania całkowitego przeżycia (*overall survival*; OS) i czasu wolnego od progresji choroby (*progression free survival*; PFS). Należało się więc spodziewać czterech krzywych ROC (OS-leptyna; OS-adiponektyna; PFS-leptyna; PFS-adiponektyna). Tymczasem na rysunku 4 przedstawiono dwie krzywe. Legenda pod rysunkiem jest bardzo lakoniczna i nie pozwala stwierdzić którą wartość, OS czy PFS, próbowano przewidzieć badając stężenia leptyny i adiponektyny. Dodatkowo w tym rozdziale przedstawiono tabelę 6. Z nagłówka tabeli wynika, że powinny się w niej znaleźć wartości stężeń mianowane w ng/ml. Tymczasem z pierwszej kolumny wynika, że podane wartości są albo procentowe, albo niemianowane, jedynie wartość odcięcia powinna być mianowana w ng/ml. Tę wartość wyznaczono za pomocą indeksu Youdena. Indeks Youdena to maksimum funkcji określonej równaniem $J = \text{czułość} + \text{swoistość} - 1$. Ponadto, w nagłówku napisano, że stężenia leptyny i adiponektyny to destymulanty, co oznacza cechy statystyczne ujemnie skorelowane ze zmienną objaśnianą. Wygląda na to, że Autorzy również w tym miejscu opisali tabelę bez zrozumienia używanych sformułowań.

W rozdziale 3.6 przedstawiono estymatory Kaplana-Meiera dla odpowiednio OS i stężenia leptyny (rysunek 5A), PFS i stężenia leptyny (5B) przy wartości odcięcia 12,38 ng/ml; OS i stężenia leptyny (rysunek 5C), PFS i stężenia leptyny (5D) przy wartości odcięcia 16,92 ng/ml; OS i stężenia adiponektyny (rysunek 6A), PFS i stężenia adiponektyny (6B) przy wartości odcięcia 27,2 ng/ml; OS i stężenia adiponektyny (rysunek 6C), PFS i stężenia adiponektyny (6D) przy wartości odcięcia 28,49 ng/ml. Na wykresach tych osie odciętych opisano jako czas od randomizacji (*Time of randomization*) mimo, że w opisanych badaniach nie występowała randomizacja. Wartości 12,38 ng/ml oraz 27,2 ng/ml to wartości środkowe (*median*) uzyskane z pomiarów leptyny i adiponektyny przed terapią, natomiast wartości odcięcia 16,92 ng/ml dla leptyny i 28,49 ng/ml adiponektyny ustalono za pomocą indeksu Youdena. Żaden z analizowanych testów nie okazał się mieć wartości predykcyjnej.

Rozdziały 3.7 oraz 3.8 zawierają analizy analogiczne do rozdziałów 3.5 i 3.6, tyle że w odniesieniu do stężeń adipokin w surowicy krwi pacjentek po zakończeniu terapii. Wszystkie błędy z rozdziałów 3.5 i 3.6 zostały powtórzone w rozdziałach 3.7 oraz 3.8. Natomiast w tych rozdziałach Autorzy podają, że badanie stężenia leptyny po zakończeniu terapii, przy wartości progowej 23,66 ng/ml lub 26,88 ng/ml może służyć do przewidywania dalszego przebiegu choroby. Nie jest jednak jasne, po co Autorzy zastosowali te dwa progi, skoro ich zastosowanie dzieliło pacjentki na dwie identyczne grupy. Autorzy najwyraźniej nie zauważyli, że pary wykresów 8A i 8C oraz 8B i 8D są takie same.

Na koniec w rozdziale 3.9 zostały przedstawione cztery modele regresji liniowej w których badano korelacje badanych stężeń z czasem wolnym od choroby (*disease free survival*; DFS). Każdy model regresji liniowej to tak naprawdę wykres funkcji, w której zmienne modyfikowane są przez przyjęte wartości parametrów. Autorzy stwierdzili, że „Modele 1, 2 i 3 niezależnie od skorygowanych parametrów wykazały, że wyższe stężenie leptyny po leczeniu korelowało z krótszym czasem wolnym od choroby ($p=0,0078$; $p=0,0031$, $p=0,0324$, odpowiednio)” (*“Model 1, 2 and 3 independently of adjusted factors showed that the higher post-treatment leptin concentration was correlated with a shorter disease-free survival ($p=0.0078$; $p=0.0031$, $p=0.0324$, respectively)”*). Te wartości p są istotne nawet po skorygowaniu ich poprawką Bonferroniego, gdyż są mniejsze niż $0.05/4$ czyli 0.0125 . Niestety funkcje dopasowujące modele do otrzymanych danych nie zostały pokazane w publikacji. Jeżeli publikuje się informację, że jakiś model może służyć do oceny przebiegu choroby, to model ten musi być ujawniony, aby inni badacze mogli z niego skorzystać i go zweryfikować.

Warto jeszcze wspomnieć, że we wprowadzeniu do publikacji 2 Autorka rozprawy doktorskiej stwierdziła (pogrubionym drukiem), że **„stężenie leptyny wyższe niż 26,88 ng/ml zwiększa prawdopodobieństwo nawrotu i śmiertelność chorobową w tej grupie”** (pacjentek z nowotworem piersi). W zdaniu tym widać dosyć częste niestety pomylenie korelacji ze związkiem przyczynowo-skutkowym. O ile taka pomyłka może być irytująca np. u dziennikarzy, to u naukowca posługującego się statystyką jest poważnym błędem.

Publikacja 3

Publikacja 3 ukazała się w roku 2023 w czasopiśmie *Biomedicines*. We wstępie do publikacji, na stronie 106 rozprawy Autorka napisała, że „celem badania była ocena wpływu markerów śródbłonkowych przed i po leczeniu [...] na LAR (oceniany również przed i po leczeniu)”. Taka ocena wpływu białek produkowanych przez śródbłonek naczyniowy na inne białka produkowane przez organizm (LAR to stosunek stężeń leptyny do adiponektyny) byłaby trudna do wykonania. Wymagałaby bardzo starannego losowego przydziału pacjentek do grup eksperymentalnych, następnie u pacjentek w poszczególnych grupach blokowania produkcji lub uwalniania poszczególnych białek produkowanych przez śródbłonek i po potwierdzeniu, że zablokowanie było skuteczne, wykonania pomiarów potrzebnych do określenia LAR. Opublikowana praca nie zawierała takich badań. Streszczenie pracy zawiera stwierdzenie, że „Celem niniejszego badania była ocena związku markerów śródbłonka (tj. sP-selektyny, sE-selektyny i czynnika von Willebranda) ze stosunkiem leptyny do adiponektyny (LAR)” (*“The aim of the present study was to assess the association of endothelial markers (i.e., sP-selectin, sE-selectin and von Willebrand factor) with the leptin-to-adiponectin ratio (LAR)”*). Ocena korelacji jest znacznie łatwiejsza do wykonania i to właśnie zostało opisane w **publikacji 3**. Niestety już we wstępie publikacji, pomyłka została powtórzona, pojawiło się tam bowiem zdanie „zatem celem niniejszego badania była ocena wpływu markerów śródbłonkowych przed i po leczeniu (tj. sP-selektyna, sE-selektyna i vWF) na LAR (również oceniane przed i po leczeniu)” (*„thus, the aim of the present study was to assess the effect of endothelial markers before and after treatment (i.e., sP-selectin, sE-selectin and vWF) on the LAR (also assessed before and after treatment)”*).

Grupę badaną w tej pracy stanowiło 70 pacjentek z inwazyjnymi nowotworami piersi, we wczesnym stadium choroby, z przerzutami najwyżej do węzłów chłonnych. U wszystkich pacjentek przeprowadzono leczenie operacyjne, a u 68 zastosowano dodatkowe leczenie, zgodnie z aktualnymi wytycznymi. U wszystkich pacjentek, przed i po zakończonym leczeniu badano stężenia leptyny, adiponektyny, rozpuszczalnej P-selektyny (*sP-selectin*), rozpuszczalnej E-selektyny (*sE-selectin*) oraz czynnika von Willebranda (*vWF*). Dodatkowo wprowadzono współczynnik LAR (*leptin-to-adiponectin ratio*), którego miarą był stosunek stężenia leptyny do stężenia adiponektyny).

Tabelę 2 w tej publikacji zatytułowano: „Charakterystyka leczenia pacjentek w odniesieniu do LAR” (*„Treatment characteristics of patients in respect of LAR”*). Tytuł ten sugeruje, że pacjentki były leczone w zależności od zmierzonego współczynnika LAR. Jednak z zawartości tabeli wynika, że porównywano współczynniki LAR przed i po leczeniu, jednakże nie w grupie wszystkich pacjentek, a u pacjentek podzielonych według trzech różnych kryteriów. W legendzie napisano, że „uznano, że wartości $p < 0,05$ wskazują na istotność statystyczną” (*„p-values < 0.05 were considered to indicate statistical significance”*). Oznacza to, że ponownie nie uwzględniono problemu porównań wielokrotnych. Tym razem na tym samym zestawie danych wykonano ich trzy, przez co za istotne statystycznie należy

uznać dopiero $p < 0,0167$. Dla czterech porównań takie wartości p zostały uzyskane, tymczasem w tabeli grubym drukiem wskazującym na istotność statystyczną, zaznaczono siedem porównań, w tym takie dla którego $p = 0,4652$.

Tabele 3, 4 oraz 5 zawierały analogiczne porównania, tyle że były w nich zamieszczone stężenia sE-selektyny (tabela 3), sP-selektyny (tabela 4) oraz vWF (tabela 5). Podobnie jak w tabeli 2 tytuły są mylące i podobnie jak w tabeli 2 nie uwzględniono problemu porównań wielokrotnych. Można wyciągnąć z tych tabel generalny wniosek, że po leczeniu wartości LAR oraz stężenia sE-selektyny, sP-selektyny i vWF rosną, w stosunku do pomiarów sprzed leczenia.

W rozdziałach 3.6, 3.7, 3.8 przedstawiono wyniki analiz, które ośmielę się nazwać „polowaniem na korelacje”. Próbowano np. korelować stężenia każdego ze zmierzonych białek produkowanych przez śródbłonek ze współczynnikami LAR (w podziale na trzy grupy) przed (tabela 6) lub po (tabela 7) leczeniu. Żadne z tych porównań nie wykazało istnienia korelacji. Nie jest też jasne, jakie miałyby być ich biologiczne znaczenie. Podobne wyniki przedstawiono za pomocą heatmapy na rysunku 2. Kolejna próba, została przedstawiona w tabeli 8. Osiem mierzonych parametrów próbowano za pomocą modeli regresji liniowej, modyfikowanych przez różne zestawy parametrów takich jak wiek, BMI, liczba potomstwa, oraz stan przed lub po menopauzie dopasować do wyników, jakimi były czasy wolne od choroby (DFS). Mimo, że żaden model nie pasował do uzyskanych wyników, to przedstawiono je w tabeli, ponownie nie ujawniając jakie funkcje stoją za tymi modelami. W legendzie tej tabeli (podobnie jak w wielu innych miejscach publikacji) umieszczono stwierdzenie: „podkreślone wartości p reprezentują bliskość istotności statystycznej” („*underlined p-values represent closeness to statistical significance*”). Podkreślone wartości p to 0,0504 oraz 0,0583. Tymczasem dla tej tabeli, z powodu zastosowania czterech modeli, za istotne statystycznie można uznać wyniki dla których $p < 0,0125$, czyli bardzo odległe od tych podkreślonych.

Dalej na rysunkach 3, 4, 5 i 6 pokazano estymatory Kaplana-Meiera dla odpowiednio: OS i stężenia sP-selektyny przed leczeniem (rysunek 3A), PFS i stężenia sP-selektyny przed leczeniem (3B) przy wartości odcięcia 265,05 ng/ml; OS i stężenia sP-selektyny przed leczeniem (rysunek 3C), PFS i stężenia sP-selektyny przed leczeniem (3D) przy wartości odcięcia 247,40 ng/ml; OS i stężenia vWF przed leczeniem (rysunek 4A), PFS i stężenia vWF przed leczeniem (4B) przy wartości odcięcia 569,90 ng/ml; OS i stężenia vWF przed leczeniem (rysunek 4C), PFS i stężenia vWF przed leczeniem (4D) przy wartości odcięcia 600 ng/ml; OS i LAR po leczeniu (rysunek 5A), PFS i LAR po leczeniu (5B) przy wartości odcięcia 0,82; OS i LAR po leczeniu (rysunek 5C), PFS i LAR po leczeniu (5D) przy wartości odcięcia 0,83; OS i stężenia sP-selektyny po leczeniu (rysunek 6A), PFS i stężenia sP-selektyny po leczeniu (6B) przy wartości odcięcia 1738,76 ng/ml; OS i stężenia sP-selektyny po leczeniu (rysunek 6C), PFS i stężenia sP-selektyny po leczeniu (6D) przy wartości odcięcia 2224,44 ng/ml. Na wykresach ponownie osie odciętych opisano jako czas od randomizacji (*Time of randomization*) mimo, że i w tych badaniach nie występowała

randomizacja. Ponownie też pokazywanie tak wielu wykresów dla bardzo podobnych wartości odcięcia jest niezrozumiałe. Znowu Autorzy nie zauważyli, że pary wykresów 5A i 5C oraz 5B i 5D są identyczne.

Należy zauważyć, że przypadku wykresów z rysunku 5A i 5C przyjęte wartości LAR (niemal identyczne) dość dobrze rozdzielały pacjentki z dobrą i złą prognozą. Gdyby przyjąć jedną z tych wartości odcięcia dla LAR i pokazać jedynie estymatory Kaplana-Meiera dla jednej z tych wartości, byłby to ciekawy wynik.

W tabeli 9 przedstawiono kolejne próby dopasowania regresji liniowej, tym razem wykonane z użyciem modelu Cox'a. Model Cox'a jest rodzajem modelu regresji zaprojektowanym specjalnie do analizy przeżycia, więc w zasadzie to ten model powinien być wykorzystany, a nie modele dla których wyniki pokazano w tabeli 8. Tym razem wykorzystano ten model do ustalenia regresji modyfikowanej wieloczynnikowo (*multivariate*) lub jednoczynnikowo (*univariate*). Czynniki wykorzystane do modyfikacji wieloczynnikowej to „BMI, wiek w chwili rozpoznania, palenie tytoniu, stopień zaawansowania, typ molekularny, typ histologiczny, przerzuty do węzłów chłonnych i wielkość guza” („*BMI, age at the time of diagnosis, smoking status, staging, molecular type, histological type, nodal metastasis, and tumor size*”). Ponieważ wykorzystano do analizy tych danych dwa modele, to za wyniki istotne statystycznie można uznać te, dla których $p < 0,025$, tymczasem w tabeli jako istotne oznaczono wartości 0,0274 oraz 0,0477. Znowu dla każdego mierzonego u pacjentek parametru zastosowano w tym modelu dwie wartości odcięcia i ponownie podano po dwa identyczne zestawy wartości dla współczynnika LAR zmierzonego po zakończeniu terapii.

W dyskusji publikacji znalazło się kilka zdań, w których najwyraźniej brakuje fragmentów, przez co są pozbawione sensu. Na przykład, na stronie 19 znalazły się następujące zdania:

- „Do dodatkowej analizy wykorzystano krzywe ROC (odpowiednio mediana i punkt odcięcia ROC) mają 10,32-krotnie większe ryzyko nawrotu choroby.” („*ROC curves were used for additional analysis (median and ROC cut-off, respectively) have a 10.32-fold higher risk of disease relapse.*”);

- „Zatem pacjentki ze stężeniem LAR przed leczeniem wyższym niż 0,46 wydają się mieć ryzyko nawrotu choroby 11,32 razy wyższe; także pacjentki z stężeniem LAR po leczeniu wyższym niż 0,82 i 0,83.” (“*Thus, subjects with a pre-treatment LAR concentration higher than 0.46 appear to have a 11.32-fold higher risk of disease recurrence; also, patients with post-treatment LAR concentrations higher than 0.82 and 0.83.*”).

Warto zauważyć, że LAR to nie jest stężenie, tylko wynik dzielenia dwóch stężeń przez siebie.

W dyskusji pojawiają się też stwierdzenia, które nie mają uzasadnienia w prezentowanych wynikach. Do takich należy na przykład: „co więcej, niższe stężenie vWF przed leczeniem jest czynnikiem prognostycznym wysokiego ryzyka wznowy raka piersi. Wynik ten potwierdzają krzywe Kaplana–Meiera oraz wieloczynnikowa i jednoczynnikowa regresja Coxa.” (“*Furthermore, a lower pre-treatment vWF concentration is a predictor of a high risk of breast cancer recurrence. This result is confirmed by the*

Kaplan–Meier curves and multivariate and univariate Cox regression.”). Podobne stwierdzenie pojawia się też w opisie rysunku 4 na stronie 12.

Należy jeszcze odnotować, że publikacja 3 w czasopiśmie zawiera dane dodatkowe (*supplementary materials*). Niestety do rozprawy doktorskiej te materiały nie zostały dołączone.

5. Dodatkowe uwagi

Nowotwory piersi są najczęściej diagnozowanymi nowotworami u kobiet. Jest to grupa nowotworów heterogenicznych, są wśród nich guzy rozwijające się bardzo agresywnie, są też guzy wolno rosnące. Najbardziej agresywne są guzy piersi rozwijające się na podłożu odziedziczonych mutacji w genach supresorowych nowotworzenia BRCA1 i BRCA2. Najczęściej stosowaną metodą diagnostyczną, mającą na celu badania przesiewowe, jest mammografia. Jest to metoda bardzo niedoskonała, gdyż często wykrywa te najbardziej agresywne nowotwory piersi na zbyt późnym etapie (np. <https://doi.org/10.3390/cancers10120477>), z drugiej zaś strony jest obarczona wysokim odsetkiem nadmiernych diagnoz (*overdiagnosis*; np. <https://doi.org/10.7326/M21-3577>, <https://doi.org/10.1136/bmj.p1823>). Dlatego jak najbardziej uzasadnione jest poszukiwanie nowych metod badań przesiewowych, łatwych do wykonania i pozbawionych powyższych wad.

Taki cel zapewne przyświecał Autorce przedstawionej do oceny rozprawy doktorskiej. Dla realizacji tego celu postanowiła zweryfikować publikowane wcześniej doniesienia dotyczące pomiarów stężenia w surowicy krwi obwodowej t-PA i PAI-1 w celu odróżniania kobiet zdrowych od chorych. We wnioskach z rozprawy Autorka twierdzi, że taki cel został osiągnięty, jednakże jako wartość stężenia PAI-1 która rozdziela grupę kobiet zdrowych od chorych podaje wartość środkową zmierzoną w grupie zdrowych (33,91 ng/ml). Są to stwierdzenia wzajemnie się wykluczające, podważające też przydatność całej publikacji. Niestety nie są to jedyne poważne błędy zauważone w **publikacji 1**.

Autorka chciała też zaproponować metody diagnostyczne pozwalające u pacjentek, u których nowotwór został zdiagnozowany tradycyjnymi metodami, przewidzieć dalszy przebieg choroby. Tymi metodami miały być pomiary stężenia w surowicy krwi obwodowej t-PA i PAI-1, leptyny i adiponektyny, sP-selektyny, sE-selektyny i vWF. Z przedstawionych badań wynika, że pomiar stężenia leptyny po leczeniu (lub stężenie leptyny/stężenia adiponektyny po leczeniu) jest najbliższy spełnienia założonego celu. Niestety ten wynik jest trudny do odnalezienia w powodzi wyników zupełnie nieistotnych, a co gorsza przedstawionych z licznymi błędami.

Dziwi, że wśród wielu parametrów, którymi opisano badane pacjentki, nie został podany stan genów BRCA-1 i BRCA-2. Nie podano, czy pacjentek z mutacjami w tych genach w ogóle nie było w analizowanych grupach (mutacje w tych genach nie były podanymi w rozprawie warunkami wykluczenia z badania), czy nie brano ich pod uwagę. Gdyby właśnie pacjentki z tymi mutacjami okazały się mieć nawroty choroby, to dodatkowy pomiar leptyny miałby niewielkie znaczenie, ponieważ fakt, że

te mutacje warunkują agresywny przebieg choroby jest dobrze udokumentowany, a te mutacje się diagnozuje.

Jak wspomniano, badania zamieszczone w rozprawie doktorskiej nie były nowatorskie. Miały potwierdzić lub zaprzeczyć badaniom opublikowanym wcześniej przez inne grupy. W dyskusjach wszystkich trzech publikacji odniesiono się do wyników wcześniejszych prac, które często były sprzeczne z wynikami tych publikacji. Rzecz w tym, że wiele wyników, które w zaprezentowanych trzech publikacjach opisano jako „istotne statystycznie” naprawdę takimi nie było. Autorzy nie biorą pod uwagę faktu, że jeśli wykona się 100 porównań statystycznych, to przy założonym poziomie istotności $p < 0,05$, 20 z tych porównań w sposób losowy trafi do przedziału „istotności”. Dlatego należy stosować narzędzia zmniejszające te losowe efekty, a ponadto nie należy mnożyć analiz poza te, które mają sens biologiczny. Jeśli również w cytowanych przez Autorów publikacji pracach analiza statystyczna była na tak niskim poziomie, nie dziwi wiele sprzecznych wyników. Uważam przy tym, że prace zawierające negatywne wyniki powinny być publikowane. Szkodliwe jest jednak szukanie „istotności statystycznej” tam, gdzie jej nie ma.

Inną kwestią, która budzi poważny niepokój jest fakt, że prace zawierające tak istotne błędy zostały zaakceptowane przez wszystkich Współautorów oraz przez recenzentów czasopism i opublikowane. Dotyczy to zarówno błędów językowych, niestarannych i niespójnych opisów, jak i poważnych błędów merytorycznych.

6. Konkluzja

Zgodnie z Art. 187. ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce:

1. Rozprawa doktorska prezentuje ogólną wiedzę teoretyczną kandydata w dyscyplinie albo dyscyplinach oraz umiejętność samodzielnego prowadzenia pracy naukowej lub artystycznej.
2. Przedmiotem rozprawy doktorskiej jest oryginalne rozwiązanie problemu naukowego, oryginalne rozwiązanie w zakresie zastosowania wyników własnych badań naukowych w sferze gospodarczej lub społecznej albo oryginalne dokonanie artystyczne.

Moim zadaniem, jako recenzenta rozprawy, było stwierdzenie czy przedstawiona do oceny rozprawa spełnia te wymogi. Jak wspomniano powyżej, rola Autorki rozprawy polegała na przeprowadzeniu analiz statystycznych wyników dostarczonych przez innych badaczy (Współautorów publikacji). Jak wynika z oświadczeń Autorki rozprawy oraz z oświadczeń Współautorów publikacji, polegała też na sformułowaniu wniosków i napisaniu prac. Niestety wszystkie te elementy są wykonane

na bardzo niskim poziomie i pełne błędów. Dlatego nie mogę stwierdzić, że Autorka posiada umiejętność samodzielnego prowadzenia pracy naukowej, w zakresie w jakim się jej podjęła.

Uważam więc, że przedstawiona do oceny praca nie spełnia wymogów określonych w art. 187 Ustawy z dnia 20 lipca 2018 roku *Prawo o szkolnictwie wyższym i nauce* (Dz. U. z 2021 poz. 478 ze zm.).

Awa Marankoska