Recenzja rozprawy doktorskiej lek. Małgorzaty Szurgot pt. „Analiza czynników rokowniczych u dzieci ze zwojakiem zarodkowym współczulnym".

Przedstawiona mi do recenzji rozprawa liczy 260 stron i ma układ typowy dla tego rodzaju dysertacji. Jest ilustrowana 161 rycinami oraz zawiera 23 tabele. Doktorantka cytuje 324 pozycje piśmiennictwa.

Doktorantka – jak wskazuje tytuł rozprawy – za swój cel obrała ocenę wartości znanych i poszukiwanie nowych czynników prognostycznych w zwojaku zarodkowym współczulnym. W tym celu poddała retrospektywnej analizie losy 180 dzieci leczonych w Klinice Onkologii i Hematologii Dziecięcej Uniwersyteckiego Szpitala Dziecięcego w Krakowie w latach 1991-2014.

Lek. Małgorzata Szurgot uwzględniła liczne parametry kliniczne (m.in. wiek, płeć, stopień zaawansowania choroby, grupa krwi), laboratoryjne (m.in. stężenia ferrytyny, katecholamin, wartość OB, aktywność LDH i NSE), histopatologiczne (typ budowy nowotworu i grupa histologiczna według INPC), molekularne (amplifikacja *MYCN*, delecja 1p, zwiększenie liczby kopii 17q, ploidia DNA) oraz dane związane z samą terapią (m.in. odpowiedź na leczenie, objawy niepożądane). Mediana czasu obserwacji klinicznej wyniosła 56 miesięcy.

Na zakończenie rozprawy Doktorantka formułuje siedem wniosków.

Uwagi ogólne

Z uznaniem należy odnieść się do wyboru tematu – zwojak zarodkowy współczulny stanowi istotny problem w onkologii dziecięcej ze względu na stosunkowo dużą częstość występowania. W tym kontekście analiza losów dzieci, które w ciągu ponad 20 lat były leczone w jednym ośrodku, ma istotne znaczenie poznawcze. Z pewnością także umożliwia wykazanie, czy osoba aspirująca do uzyskania stopnia naukowego doktora potrafi prawidłowo prowadzić pracę badawczą.

Praca napisana jest starannie, komunikatywnym językiem. Błędy literowe są nieliczne chociaż jeden z nich znalazł się nawet w spisie treści (punkt 1.3). Stylistyka jest mocną stroną dysertacji, chociaż zdarza się Autorce używać sformułowań w rodzaju „leczenie chemioterapią" co stanowi oczywisty i rażący pleonazm albo niezręczności w rodzaju „mniejsze przeżycie". Są to jednak wyjątki.

Zastrzeżenia mam do sposobu pisania nazw genów – zgodnie z przyjętymi także i w naszym kraju zasadami, nazwy genów człowieka powinny być pisane wielkimi literami i kursywą. Wskazane byłoby także konsekwentne uwzględnienie oficjalnych skrótów ich nazw. Autorka na przykład wymiennie używa nazw „NMYC" i „MYCN", ani razu nie stosując kursywy, a prawidłowa pisownia skrótu nazwy tego genu to *MYCN*.

Z przyjemnością natomiast zauważyłem częste używanie spolszczonych wersji międzynarodowych nazw leków cytotoksycznych. Niestety dostrzegłem tu wyjątki, np. w części Wstępu dotyczącej leków molekularnych Autorka niekiedy wymienia nazwy angielskie *„crizotinib"* i do tego zdarza się Jej używać wtedy polskiej deklinacji, np. *„palbociclibu"*, a na stronie 60 pojawia się „temodal" zamiast temozolomidu. Pierwszy z tych wyjątków byłby zrozumiały w odniesieniu do leków, które w chwili pisania rozprawy nie miały jeszcze polskich nazw (oczywiście pisząc je w mianowniku), ale obydwa wymienione preparaty są znane od dawna jako palbocyklib i kryzotynib. Z kolei nazwa irynotekan pisana jest albo poprawnie, albo niepoprawnie „irinotekan".

Staranność edytorską rozprawy ogólnie oceniam jednak jako bardzo dobrą.

W tekście rozprawy nie znalazłem informacji o uzyskaniu zgody Komisji Bioetycznej na przeprowadzenie badania. Pomimo, że badanie ma charakter retrospektywny, nie wiąże się z jakimkolwiek ryzykiem dla pacjentów i zgoda Komisji Bioetycznej wydaje się być czystą formalnością, informacja o uzyskaniu takiej zgody powinna znaleźć się w tekście rozprawy. **Zakładam oczywiście, że taką zgodę uzyskano. Jeśli nie, rozprawa w ogóle nie powinna być poddana recenzji, ponieważ wtedy nie spełniono by podstawowego kryterium umożliwiającego prowadzenie jakichkolwiek badań naukowych na ludziach.**

Ocena poszczególnych części rozprawy

Wstęp

W liczącym sobie niemal 50 stron Wstępie Autorka bardzo obszernie omawia problematykę związaną ze zwojakiem zarodkowym współczulnym poczynając od epidemiologii, poprzez symptomatologię, diagnostykę, klasyfikację stopnia zaawansowania, patomorfologię, biologię molekularną, czynniki rokownicze, aż do leczenia. Tematyka bezpośrednio związana z celem pracy zajmuje 12 stron, czyli 1/4 objętości. Nie rozumiem celowości szczegółowego omawiania np. symptomatologii (aż 4 strony). Jestem pełen

podziwu dla systematyczności i pracowitości Autorki, ale uważam, że we Wstępie zostały zaburzone proporcje ilościowe między informacjami rzeczywiście niezbędnymi dla zrozumienia dalszej części rozprawy, a efektami chęci tworzenia monografii dotyczącej omawianego nowotworu. To właśnie spowodowało, że Wstęp jest zbyt długi i zawiera wiele – moim zdaniem - niepotrzebnych informacji.

Materiał i metody

Autorka zawarła w tekście informację o tym, że porównania między krzywymi przeżycia Kaplana-Meiera dla więcej niż dwóch grup wykonano w oparciu o test log-rank. Pomijając fakt, że stwierdzenie, że test log-rank porównuje krzywe przeżycia jest trochę nieprecyzyjne, bo test *de facto* porównuje liczby wydarzeń między grupami, a krzywe są tylko graficzną ilustracją funkcji przeżycia, to intryguje mnie jakiego mianowicie testu używano do porównań między dwiema grupami, bo takich porównań jest bardzo dużo.

Kolejna istotna niejasność to posługiwanie się czasem wolnym od niekorzystnych zdarzeń (EFS) bez podania definicji, czym są te wydarzenia. Zarówno DFS i OS mają jednoznaczne definicje, jednak definicja EFS nie ma charakteru uniwersalnego.

Wreszcie, Autorka uznała, że p większe od 0,05, ale mniejsze od 0,1 świadczy o tendencji statystycznej, a p mniejsze od 0,05 o istotności. Powstaje zatem przewrotne pytanie, o czym w takim razie według Doktorantki świadczy p równe 0,05?

Wyniki

Ta część jest bardzo obszerna i zajmuje niemal 120 stron. Moim zdaniem konsekwentne ilustrowanie rycinami niemal wszystkich badanych czynników nie jest potrzebne, a przyczynia się do niepotrzebnego zwiększenia objętości. Przykładem może być wiek, którego rozkład ilustrowano aż 6 rycinami (strony 65-67). Z kolei wpływ wieku na czas przeżycia ilustrowano 5 rycinami (strony 71-74) dychotomizując wiek pacjentów na 5 sposobów. Powstaje pytanie, dlaczego dychotomizowano w taki właśnie sposób? Dychotomizacja danych ciągłych jest procesem wiążącym się z utratą części informacji dlatego szkoda, że nie wykonano testu oceniającego wpływ na rokowanie wieku traktowanego jako zmienna ciągła. Ta sama uwaga dotyczy wszystkich badanych czynników rokowniczych mających charakter zmiennych ciągłych.

Na stronie 69 Autorka podaje, że średni czas przeżycia wolnego od objawów choroby dotyczy 179 dzieci i uwzględnia pacjentów z odpowiedzią na leczenie i bez odpowiedzi na leczenie oraz że wynosi 61,6 mies. Nie wiem, co Doktorantka ma na myśli pisząc o przeżyciu wolnym od objawów choroby. W metodyce nie znalazłem na ten temat wzmianki. Z pewnością nie jest to czas przeżycia wolnego od choroby (DFS).

Nie wiem, jak był wyliczany średni czas przeżycia. Czy uwzględniano estymację metodą Kaplana-Meiera czy też to prosta średnia arytmetyczna rzeczywiście obserwowanych czasów przeżycia? To samo dotyczy mediany przeżycia. Tu mam poważne podejrzenie, że nie uwzględniano estymatora Kaplana-Meiera, ponieważ na stronie 71 znajduje się wykres ilustrujący OS w całej grupie i po 5 latach odsetek chorych żyjących wynosi 69% (pisze o tym Autorka na stronie 69). Trudno spodziewać się, że przy takim przebiegu krzywej ilustrującej funkcję przeżycia, mediana czasu przeżycia wynosi tylko 56 miesięcy. Jeśli tak, to taką metodykę należy uznać za bardzo niestandardową.

Na stronie 75 Autorka pisze, że objawów choroby nie było u 11 dzieci, ale w nawiasie jest błędnie podany odsetek procentowy (61,3%).

Nie wiem, dlaczego Autorka podaje – omawiając różne czynniki – różną liczebność grupy badanej. Na przykład, na stronie 98 zawarto informację, że morfologia krwi była prawidłowa u 76 dzieci, a nieprawidłowa u 97 pacjentów. Czy mam przez to rozumieć, że u 7 dzieci nie wykonano w ogóle morfologii? Czy też, że ją wykonano, ale wynik zaginął? Autorka omawiając na stronie 101 znaczenie aktywności LDH napisała, że oznaczono ją u 166 ze 173 pacjentów, a na stronie 103 napisała, że aktywność NSE oznaczono u 97 ze 172 pacjentów. Z kolei na stronie 107 wspomniała, że dobową zbiórkę moczu przeprowadzono u 170 pacjentów ze 177. Na stronie 121 stwierdza, że badanie histopatologiczne wykonano u 126 ze 179 pacjentów. Przypomnę, że grupa badana liczyła 180 dzieci, a nie 173, 172, 177 czy 179.

Intryguje mnie, jak Doktorantce udało się przeprowadzenie analizy przeżycia przy pomocy estymatora Kaplana-Meiera i zastosowanie testu log-rank w niekiedy bardzo małych i wybitnie niezrównoważonych liczebnie grupach z nielicznymi obserwacjami kompletnymi. Skrajnym przykładem ilustrującym moje wątpliwości jest analiza przedstawiona na stronie 95 wskazująca na brak istotnej statystycznie różnicy (p=0,43424) pomiędzy OS 3 (słownie:

trzech) pacjentów z guzami wytwarzającymi VIP (wszystkie obserwacje cenzurowane - nie było zgonów) i OS pozostałych 177 pacjentów.

Począwszy od strony 107 Doktorantka omawia znaczenie katecholamin i ich metabolitów. Znalazłem tu niestety nieścisłości. Autorka stwierdza – jak już wspomniałem wyżej - że u 170 pacjentów przeprowadzono dobową zbiórkę moczu na aminy katecholowe i ich metabolity. Zdanie dalej twierdzi, że u 151 na 154 osób guz wydzielał katecholaminy, a nieco dalej precyzuje, że u 3 nie wydzielał. Powstaje pytanie, czy katecholaminy oznaczono u 154 czy też u 170 pacjentów? Omawiając dalej poszczególne katecholaminy Doktorantka stwierdza, że stężenie adrenaliny było podwyższone u 7 osób, a prawidłowe u 138. Oznacza to, że oznaczono adrenalinę u 145 osób. Z kolei noradrenalinę oznaczono u 146 osób, a dopaminę u 143 osób itd. Nie znalazłem w żadnym przypadku oznaczenia jakiejkolwiek katecholaminy lub metabolitu u 170 osób.

Kiedy zestawić powyższe dane z wynikami dotyczącymi różnic w aktywności enzymów i stężeń katecholamin w zależności od obecności lub nieobecności przerzutów odległych, wątpliwości pogłębiają się. Na stronie 101 zawarto informację, że aktywność LDH oznaczono u 166 chorych, a z Ryc. 58 umieszczonej na tej samej stronie wynika, że u 163. Analizując Ryc. 79 dochodzę do wniosku, że liczba 163 jest być może prawdziwa. Z kolei na stronie 117 wspomniano, że aktywność NSE oznaczono u 40 osób z przerzutami i u 51 osób bez przerzutów odległych (czyli łącznie u 91 osób). Z kolei wcześniej, na stronie 103 Autorka twierdzi, że NSE oznaczono u 97 osób. Czy wytłumaczeniem jest, że u części osób nie było danych dotyczących obecności lub nie przerzutów odległych? Jak wobec tego wytłumaczyć to, że stopień zaawansowania określono u wszystkich 180 osób (Ryc. 8)?

Na stronie 142 Autorka pisze, że oceny węzłów chłonnych dokonano u 107 pacjentów i u 86 z nich były powiększone, a u 21 - niepowiększone. Zaraz potem stwierdza, że u 31 pacjentów z 64 (48%) miały wielkość co najmniej 2 cm. Skąd wzięła się liczba 64?

Podałem wyżej tylko przykłady nieścisłości danych dotyczących liczebności grup. Nie są to duże rozbieżności, ale niewątpliwie nie powinno ich być w ogóle, ponieważ ich występowanie w takiej skali (nieścisłości są regułą, a nie wyjątkiem) podważa wiarygodność uzyskanych wyników.

Ryc. 128 stanowi dla mnie zagadkę – co Autorka chciała wyrazić poprzez wykres funkcji przeżycia z obserwacjami kompletnymi i cenzurowanymi umieszczając na osi pionowej prawdopodobieństwo przeżycia, a na osi odciętych czas od rozpoznania do rozpoczęcia leczenia? Podobnie niezrozumiałe są dla mnie Ryc. 141, 142, 147 i 148.

Jaki był klucz doboru parametrów poddanych analizie wieloczynnikowej? Liczba czynników istotnych w analizie jednoczynnikowej jest ogromna, a uwzględniono tylko część z nich i to w wybranych arbitralnie dychotomizacjach. W analizie wieloczynnikowej uwzględniono m.in. płeć, a dla tego parametru nie znalazłem wyniku analizy jednoczynnikowej. Doktorantka poddała analizie parametry, które częściowo się pokrywają, np. stopień zaawansowania i obecność przerzutów odległych. Rozumiem, że dokonywano dwóch analiz – jedna dotyczyła czynników klinicznych, a druga - laboratoryjnych. Ciekawe, że grupa prognostyczna nie miała znaczenia.

Autorka konsekwentnie podaje wartości p z dokładnością do 5 miejsc po przecinku – taka dokładność jest niepotrzebna.

Dyskusja

W znacznej części dyskusja jest niepotrzebnym powtórzeniem uzyskanych wyników, łącznie z podaniem wartości p. Zawarto w niej nawet krótką część dotyczącą ogólnej charakterystyki pacjentów. Niestety komentarze uzyskanych wyników w kontekście danych z piśmiennictwa albo ich własna interpretacja, stanowią objętościowo mniejszość.

Wiele z uzyskanych wyników wymaga omówienia. Na przykład, na stronach 168-169 znalazła się skądinąd trywialna informacja, że chorzy, u których wystąpiła wznowa mieli krótszy czas przeżycia. Intrygującą mnie informacją jest ta, że w grupie bez wznowy choroby zmarło aż 18% osób. Czy mam rozumieć, że niemal co piąty pacjent w grupie bez nawrotu zmarł z innych przyczyn niż nowotwór? Wydaje mi się to mało prawdopodobne. Ponadto z informacji na stronie 167 oraz Tabeli na stronie 251 wynika, że całkowitą remisję uzyskano u 141 osób. Jak zatem definiowano wznowę lub jej brak, skoro wznowa wystąpiła u 42 dzieci, a nie wystąpiła u 130? Przy okazji, czy dla 8 dzieci nie udało się uzyskać informacji o wznowie?

Na stronie 158 zawarto informację, że chorzy, u których w trakcie resekcji guza usunięto całkowicie okoliczne węzły chłonne mieli gorsze rokowanie w porównaniu z tymi, u których węzłów nie usuwano w ogóle lub usunięto tylko częściowo. Tego rodzaju informacja

bez uzupełnienia jej o dane dotyczące tego, czy limfadenektomia była selektywna czy też elektywna i w jaki stopniu usunięte węzły były zajęte przerzutami, jest mało wartościowa. Ważne byłoby także uwzględnienie rozkładu innych czynników rokowniczych. W dyskusji widziałbym miejsce właśnie do takich rozważań. Wprawdzie w dyskusji Doktorantka poświęciła temu wyjątkowo dużo miejsca, bo prawie dwie strony, ale skupiła się na omówieniu mechanizmów odpowiedzi immunologicznej, a nie na wspomnianych rozważaniach. Na szczęście nie sposób się nie zgodzić z konkluzją Autorki, że wymaga to dalszych badań.

Warto byłoby w tej części omówić różnice między czynnikami rokowniczymi, a predykcyjnymi i wskazać na praktyczne trudności w odróżnieniu jednych od drugich w odniesieniu do pacjentów poddawanych terapii, kiedy oceniany jest czas całkowitego przeżycia.

Wnioski

Stanowią w większości raczej podsumowanie wyników, a nie rzeczywiste wnioski. Do tego niejasne są dla mnie kryteria wyboru najważniejszych czynników rokowniczych uwzględnionych w tej części. Na przykład, niedokrwistość jest wymieniona w pierwszym wniosku jako niekorzystny czynnik rokowniczy, a nie znalazłem niedokrwistości w analizie wieloczynnikowej. Autorka nie wspomniała o żadnym z czynników istotnych w analizie wieloczynnikowej.

Widziałbym w tym miejscu wskazanie na przykład na to, czy wyniki leczenia omawianych pacjentów są podobne jak w innych ośrodkach albo czy zmieniły się na korzyść w czasie albo wreszcie czy pacjenci w trakcie rozpoznania choroby mieli w całym okresie podobny rozkład czynników rokowniczych.

Ciekawi mnie także, które z uzyskanych wyników stały się podstawą do sformułowania wniosku nr 5.

Podsumowanie recenzji

W niniejszej recenzji zawarłem wiele krytycznych uwag i wskazałem na wiele niejasności. Niewątpliwie lektura Dysertacji nie jest łatwa i to wcale nie z uwagi na jej objętość, ale przedstawienie olbrzymiej liczby danych, których spora część sprawia wrażenie, że Autorka bardzo mechanicznie podeszła do zagadnienia sprawdzając wszystko, co tylko

znalazła. Ta ilość danych przyczynia się do tego, że powstają powtórzenia i nieścisłości liczbowe.

Nie zmienia to jednak faktu, że Doktorantka wykazała się wielkim zaangażowaniem, systematycznością i robiącą ogromne wrażenie pracowitością. Ilość zebranych i opracowanych danych jest imponująca. Moim zdaniem to właśnie te cechy w połączeniu z dużą wiedzą Doktorantki stanowią zasadniczą wartość Rozprawy.

Długo zastanawiałem się jaki powinien być końcowy wniosek mojej recenzji. Ustawa z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki w Art. 13 punkt 1 stanowi że: „Rozprawa doktorska, przygotowywana pod opieką promotora albo pod opieką promotora i promotora pomocniczego, powinna stanowić oryginalne rozwiązanie problemu naukowego lub oryginalne rozwiązanie problemu w oparciu o opracowanie projektowe, konstrukcyjne, technologiczne, lub oryginalne dokonanie artystyczne, oraz wykazywać ogólną wiedzę teoretyczną kandydata w danej dyscyplinie naukowej lub artystycznej oraz umiejętność samodzielnego prowadzenia pracy naukowej lub artystycznej." Ocenię zatem tę rozprawę ściśle w kontekście spełniania wymogów ustawowych. Niewątpliwie lek. Małgorzata Szurgot wybrała jako temat pracy adekwatny problem naukowy i dokonała jego oryginalnego rozwiązania. Bez żadnej wątpliwości – po lekturze Wstępu i w nieco mniejszym stopniu Dyskusji – mogę także stwierdzić, że Doktorantka dysponuje odpowiednią ogólną wiedzą teoretyczną w reprezentowanej dziedzinie. Ostatnim elementem jest umiejętność samodzielnego prowadzenia pracy naukowej. Zbieranie danych klinicznych, ich analiza, a także formułowanie wniosków są istotnymi elementami pracy naukowej. Przedstawiłem swoje zastrzeżenia dotyczące poprawności części przyjętej metodyki, staranności analizy danych, a także sposobu interpretacji wyników, ale łącznie uważam, że Doktorantka wykazała się jednak umiejętnościami, o których mowa w Ustawie.

Na zakończenie zatem stwierdzam, że przedłożona mi do recenzji praca – mimo bardzo licznych niedostatków - spełnia jednak wymagania ustawowe stawiane rozprawom doktorskim. W związku z tym wnioskuję o dopuszczenie lek. Małgorzaty Szurgot do publicznej obrony. Warunkiem jest jednak, że Doktorantka uzyskała zgodę Komisji Bioetycznej na przeprowadzenie badania.

Łódź, 7 maja 2020 r.

KIEROWNIK
Oddziału Chemioterapii Nowotworów
z Pododdziałem Chemioterapii Jednego Dnia
Klinika Chemioterapii Nowotworów
Prof. dr hab. n. med. Piotr Potemski