Ocena pracy doktorskiej
mgr Łukasza Kaczyńskiego
pt. „Ocena efektywności praktycznej jako kluczowego elementu raportów oceny technologii medycznych. Pragmatyczne badanie kliniczne"

Pragmatyczne badania kliniczne prowadzi się aby zebrać dane potrzebne do planowania działań w celu optymalizacji opieki zdrowotnej. Badania pragmatyczne najczęściej analizują koszty dotychczasowej praktyki klinicznej w porównaniu z efektywnością kliniczną nowej badanej interwencji. Wyniki takich analiz są pomocne w wyborze najskuteczniejszych interwencji terapeutycznych, zwłaszcza w krajach o dużym udziale w rynku prywatnych podmiotów. Ponadto badania pragmatyczne analizują wyniki świadczeń opieki zdrowotnej, uwzględniając preferencje pacjentów i koszty, w celu wykorzystania ich w procesie decyzyjnym w ochronie zdrowia.

Rozprawa doktorska mgr Łukasza Kaczyńskiego, jak można sądzić na podstawie tytułu, miała pokazać znaczenie oceny efektywności praktycznej w raportach oceny technologii medycznych, a podtytuł „pragmatyczne badanie kliniczne" mógłby sugerować, że Doktorant podjął się ambitnego zadania przeprowadzenia takiego właśnie badania. Jednak dalsza lektura rozprawy pokazuje nieco inne jej cele. Celem rozprawy, jak to sprecyzował Doktorant, było: "odnalezienie i usystematyzowanie informacji na temat pragmatycznych badań z randomizacją" oraz „ich ujednolicenie w postaci jednego kwestionariusza". Cel sformułowano niezwykle szeroko i zarazem ogólnikowo. Nie jest jasne użycie słowa „ujednolicenie" w kontekście informacji na temat badań. Podobnie tajemniczo brzmi jako cel oryginalnej rozprawy doktorskiej „odnalezienie informacji". W mojej ocenie nowy badawczy aspekt takiego celu rozprawy stanowi opracowanie autorskiego kwestionariusza, który jak można wnioskować z dalszych części tekstu mógłby służyć analizie badań klinicznych bez randomizacji. Wstęp przygotowany przez Doktoranta wprowadza kluczowe zagadnienia dotyczące nie tylko pragmatycznych badań klinicznych. Zwięźle Doktorant przypominał podstawowe pojęcia EBM, badania z randomizacją i bez niej, przeglądy systematyczne oraz zasady oceny technologii medycznych.

Jak zawsze w ocenie rozprawy w jej tradycyjnej formie, nie poddanej wcześniej zewnętrznej recenzji eksperckiej, najważniejsza jest metodologia. Opis metodyki mgr Kaczyński zawarł na zaledwie 4 stronach rozprawy. Doktorant podał na s. 15, że realizował swoje cele w oparciu o

wytyczne AOTMiT z 2016 roku, rozporządzenia Ministra Zdrowia z 02.04.2012 r. oraz wytyczne dotyczące prowadzenia przeglądów systematycznych Cochrane Collaboration (wersja 5.2.0) podając odpowiednie pozycje piśmiennictwa. Następnie Doktorant dość nieprecyzyjnie zaznaczył, że wykonywał „przeglądy medycznych baz danych w tym systematyczne" dodając, że te bazy to MEDLINE, EMBASE i/lub CENTRAL; czas ostatniego przeszukania zaznaczono dla 2 pierwszych baz i był nim 31.12.2018 r. Użycie „i/lub" nie jest jasne i brak precyzji co do przeszukiwanych baz danych jest ważną wadą tak przedstawionej metodyki. Doktorant w metodologii powołuje się na wytyczne przygotowania przeglądów systematycznych (Cochrane), ale przedstawiony przez Doktoranta opis i metodologia są dalekie od tych wytycznych.

Wyniki (rozdział 4) zawarto na ponad 100 stronach. W pierwszym akapicie Doktorant podał, że „w opisie uwzględniono 93 publikacje". Jednak kryterium wcześniej podane czyli zawartość „istotnych informacji z zakresu teoretycznych zagadnień na temat pRCTs, jak i o narzędziach służących do ich właściwego projektowania oraz oceny jakości" jest zbyt ogólne i nie jest zgodne z aktualną metodologią. Niejasne są zatem kryteria doboru badań do tej części pracy oraz kryteria uznania opisu w danej publikacji za dotyczące badania pragmatycznego. Ten brak precyzji jest dużym metodycznym ograniczeniem i nie pozwala powtórzyć wyszukiwania badań z baz danych z tym samym rezultatem na podstawie przedstawionej metodologii, co stanowi jedno z kryteriów poprawnego przeglądu systematycznego. W rozdziale 4 pojawiają się słowa kluczowe np. na s. 29 choć w tym samym zdaniu sformułowanie „szybki przegląd" nie kojarzy się z wnikliwą oceną typową dla rzetelnego badania naukowego. Znaczna część rozdziału 4 od strony 23 do 84 zawiera opis badań z randomizacją i porównanie ich z tymi pragmatycznymi z wykorzystaniem licznych pozycji piśmiennictwa. W mojej ocenie większość rozdziału 4 pt. Wyniki Wyszukiwania (od s. 23) jest właściwie rozbudowanym wstępem, przedstawieniem stanu wiedzy na temat tego rodzaju badań, ich definicji, zagadnień etycznych i prawnych oraz zasad oceny ich jakości, a także kwestionariusza CONSORT, PRECIS (także PRECIS-2), ASPECT-R. Większość tego rozdziału może być traktowana jako artykuł poglądowy.

Niejasne jest w jaki sposób dokonano wyboru przedstawionych w tabeli 4 na s. 30 cech metodologicznych (nie podano referencji w tabeli) ani według jakich kryteriów je wybrano.

Trudno powiedzieć jak do poprzednich części ma się rozdział 4.1.9. cytujący uwagi o aktualizacji wytycznych AOTiM bez podania źródła i autora.

Na stronach od 90 do 121 Doktorant opisał narzędzia do oceny badań nadając tej części cechy typowe dla przeglądu systematycznego. Jednak metodologia wybrana nie jest spójna z tym zamysłem. Wiele kwestii budzi zastrzeżenia w tym zakresie wskazując na nieścisłości, a taki brak precyzji nie jest zgodny z samą filozofią współczesnej epidemiologii klinicznej opartej na zasadach EBM.

Strategia wyszukiwania przedstawiona w rozprawie obejmowała pojedyncze słowa kluczowe bez synonimów, bez informacji jakiego rodzaju słowa kluczowe zostały użyte, i w której bazie. Można wyrazić żal, że Doktorant najpewniej nie sprawdził podstawowych źródeł informacji w tym zakresie zawierających wiele skal, np. strona Equator network, strona Centre for Evidence based medicine. Ważną pozycją piśmiennictwa pomocną w podjętych działaniach mogłaby być praca Deeksa i współpracowników (Deeks J, Dinnes J, D'Amico R, Sowden A, Sakarovitch C, et al. Evaluating non-randomised intervention studies. Health Technol Assess 2003;7(27)), która szczegółowo analizowała 194 skale używane do oceny badań bez randomizacji wraz z porównaniem domen oraz analizami metodologicznymi.

Nie podano informacji ile badań zostało wykluczonych na poszczególnych etapach pomimo opisania tej części rozprawy jako przeglądu systematycznego i przywoływania przez Doktoranta narzędzi do raportowania różnych rodzajów badań. Szkoda, że Doktorant nie przedstawił swojego badania zgodnie z przyjętymi obecnie zasadami.

Podane w rozdziale 4.2.4 kryteria włączenia skal są niejasne. Podobne wątpliwości budzi wybór skal opisanych w rozdziale 4.2.5. Niejasne jest dlaczego Doktorant przywołał skalę AXIS do badań przekrojowych, a pominął inne, dostępne narzędzia do tego typu badań (np. z NIH)?

Dlaczego Doktorant uwzględnił jedno narzędzie z JBI dla badań oceniających prevalence (które znajduje się na stronach JBI), ale pominął narzędzie JBI dla badań kohortowych? Dlaczego wymieniono kwestionariusze STARD i STROBE służące do raportowania badań diagnostycznych a nie ma innych, podobnych narzędzi do raportowania innych rodzajów badań, np. RECORD lub TRIPOD? Dlaczego opisano narzędzie QUIPS, a nie ma innych narzędzi do badań diagnostycznych i prognostycznych, zwłaszcza QUADAS? Jednocześnie Doktorant uwzględnił narzędzie ROBINS-E, które przez samych jego autorów zostało określone jako wstępne (preliminary) i nie zostało jeszcze opublikowane (a było to jedno z kryteriów włączenia), a cytowanie podane przez Doktoranta jest odwołaniem do artykułu będącego krytyką tej wstępnej wersji. Sugeruje to możliwy błąd w wyszukiwaniu i niejednoznaczne kryteria włączenia badań. Nawet jeśli Doktorant mógł arbitralnie wybrać tylko

pewne narzędzia, to przynajmniej w rozdziale Dyskusja, który zwykle dowodzi dobrej orientacji Doktoranta w temacie jego dysertacji, inne niewykorzystane narzędzia powinny zostać omówione zwłaszcza jeśli Doktorant wprowadził nowe narzędzie i uznał je za bardziej przydatne od wcześniej używanych. W rozdziale 4 najwięcej miejsca Doktorant poświęcił opisaniu poszczególnych narzędzi w oparciu o piśmiennictwo, co w mojej ocenie trudno nazwać „wynikami". Niejasne są kryteria dokonanej przez Doktoranta krytycznej oceny opisanych narzędzi (rozdział 4.2.6), a ich omówienie zawarte w rozdziale 4 pt. Wyniki Wyszukiwania jest, w mojej ocenie, w dużej mierze subiektywne i niepoparte pogłębioną analizą ani własnymi wynikami. Rozdział jest właściwie fragmentem poglądowego opracowania wybranych narzędzi.

W rozdziale 4.2.7 Doktorant przedstawił nowy kwestionariusz SQAROS jako narzędzie oceny i ta część zasługuje na szczególną uwagę w pracy doktorskiej. Jednak w mojej ocenie Doktorant mało dokładnie omówił w jaki sposób to narzędzie powstało, jak wybrano domeny i pytania oraz przyporządkowano punktację. Jednak najważniejszą wadą nowego kwestionariusza Doktoranta jest brak odpowiedniej walidacji opracowanego narzędzia. Użyto co prawda tego narzędzia do oceny badań, ale nie porównano właściwie wyników tej oceny z inną znaną skalą.

Doktorant postanowił przeprowadzić przegląd systematyczny porównujący skuteczność olanzapiny z arypiprazolem u chorych na schizofrenię, co opisano począwszy od s. 122, gdzie rozpoczyna się rozdział 4.3 pt. Zastosowanie Odnalezionych/Opracowanych Narzędzi. Po analizie baz PubMed i EMBASE do dnia 14.09.2019 r. pod kątem badań z randomizacją na temat jednorazowego podania olanzepiny w porównaniu z arypipazolem w postaci doustnej lub domięśniowej u chorych ze schizofrenią wraz z badaniami bez próby kontrolnej dotyczących samej planzepiny, Doktorant zidentyfikował 353 „rekordów", a ostatecznie10 pełnotekstowych artykułów i 3 dodatkowe publikacje, dodając diagram QUORUM. W typowym przeglądzie systematycznym powinno się przedstawić teoretyczne przesłanki do podjęcia akurat tego tematu, czyli w jakim kontekście klinicznym na pytanie kliniczne ma odpowiedzieć przygotowywany przegląd. Doktorant opisał przegląd dotyczący oceny skuteczności i efektywności olanzapiny jako systematyczny, jednakże nie przedstawił opisu ani wyników w sposób zgodny z metodologią takiego rodzaju badania. Jeśli uwzględni się zagadnienia zawarte w narzędziu AMSTAR 1 (narzędziu rekomendowanym w wytycznych AOTMiT) to w przypadku przedstawionej analizy można wyróżnić następujące zastrzeżenia:

brak jednoznacznej informacji o protokole, selekcji badań, która powinna być wykonana przez 2 osoby niezależnie,

niewystarczające wyszukiwanie – 3 bazy, bez odniesienia się do problemu wyszukiwania badań niepublikowanych (np. dostępnych u producenta leku) i ograniczenia włączenia badań ze względu na język, co może skutkować błędem systematycznym.

Strategia wyszukiwania badań do tego przeglądu zawierała pojedyncze słowa kluczowe dla każdego z elementów pytania badawczego w postaci „free textwords", a według zasad tworzenia takiej strategii powinna ona obejmować jak najwięcej synonimów oraz powinna obejmować słowa kluczowe z Tezaurusa odpowiedniego dla danej bazy (np. MESH, EMTREE). Nie opisano, w której bazie zastosowano podaną strategię. Pomimo opisania, że kwalifikowano badania dotyczące samej olanzapiny użyta strategia, w mojej ocenie, umożliwiała znalezienie badań, w których występowały oba słowa kluczowe. Nie uwzględniono listy badań wykluczonych wraz z przyczynami ich wykluczenia. Warto także zaznaczyć, że diagram QUOROM przytoczony przez Doktoranta został w 2009 roku zaktualizowany i zmienił nazwę na PRISMA.

Dyskusja w rozprawie doktorskiej jest bardzo krótka (s. 135-138) i nie odnosi się do wszystkich ograniczeń. Początek nie wydaje się odnosić do wcześniejszych rozdziałów, ale stanowiąc kolejny zbiór uwag dotyczących prób klinicznych z randomizacją i niewystarczającej popularności narzędzia PRECIS-2. Doktorant podkreślił nowy aspekt rozprawy tj. opracowanie "nowego kwestionariusza oceny jakości badań obserwacyjnych SQUAROS". Zdaniem Doktoranta nowe narzędzie „wykazuje pewną przewagę" nad obecnie rekomendowanymi narzędziami NICE i NOS, przedstawiając argumenty na korzyść swojego opracowania, w tym jego prostotę zarzucając „nadmierne rozbudowanie" wcześniejszym narzędziom. Jednak krytyka uznanych narzędzi nie została w zadawalający sposób poparta danymi porównawczymi, w tym odpowiednią walidacją. Spekulacja na temat wyższości własnej skali jest wyrazem opinii Doktoranta i moim zdaniem nie jest oparta na wynikach własnych badań. Nazwanie swojej skali efektywną z powodu krótkiego czasu (10 minut) jaki zajmuje jej wykorzystanie w ocenie badania nie jest właściwe; szybkość nie jest synonimem efektywności. Zaskoczenie powoduje zdanie pierwsze z ostatniego paragrafu dyskusji, w którym Doktorant stwierdził, że nie wykonano przeglądu systematycznego prac na temat pRCT, wybierając „najbardziej wiarygodne i ukazujące najistotniejsze informacje publikacje". Tak subiektywne podejście do doboru badań przy niespełnieniu zasad opracowania systematycznego przeglądu mimo odwoływania się wcześniej w tekście do wymagających zaleceń metodycznych wprowadzonych przez Cochrane Collaboration podważa wartość rozprawy, ograniczając wartość nowej wiedzy, która powinna być owocem rozprawy doktorskiej. Znajomość

metodologii badań także jest podważona przez takie końcowe uwagi. Pięć wniosków na s. 139-140 nie zawiera nowych obserwacji a podsumowuje informacje znane wśród osób zajmujących się metodologią badań klinicznych, a wniosek 4 nie jest poparty danymi. Doktorant nazwał „skutecznym wykorzystaniem" SQAROS przeprowadzenie przeglądu. Nie tak ocenia się nowe kwestionariusze.

Pewne sformułowania użyte przez Doktoranta są dla mnie niejasne m.in. „kwestionariusz PRECIS-2 nie ma zastąpić oceny ryzyka błędu systematycznego w pragmatycznych badaniach klinicznych z randomizacją na podstawie Cochrane Collaboration" (s. 83) – co Doktorant miał na myśli używając terminu na podstawie Cochrane Collaboration w tym zdaniu?

W podrozdziale 4.2 pt. Badania Obserwacyjne, na s. 90 pojawia się zdanie "Celem odnalezienia narzędzi służących do oceny jakości badań klinicznych bez randomizacji przeprowadzono przegląd systematyczny…", a dalej „Do analizy w pełnym tekście włączono te publikacje, które prezentowały bezpośrednio opis służących do oceny jakości badań obserwacyjnych…". Czy rzeczywiście Doktorant traktuje jako synonimy 2 różne rodzaje badań tj. interwencyjne bez randomizacji i obserwacyjne (s. 90)?

Podsumowując, mgr Łukasz Kaczyński podjął się ciekawego zadania opracowania nowego narzędzia do oceny badań bez randomizacji. Rezultat ma potencjał, jednak nie przedstawiono przekonywujących wyników potwierdzających wartość nowego narzędzia, a przegląd przygotowany nie jest zgodny z aktualnymi zaleceniami dotyczącymi takich analiz. Liczne nieścisłości i braki przedstawione powyżej uniemożliwiają pozytywną ocenę przedstawionej rozprawy doktorskiej. Powyższe uwagi przedstawiam Radzie Dyscypliny Nauki Medyczne CM UJ jako argumenty do podjęcia ostatecznej decyzji w sprawie nadania lub nie stopnia doktora Panu Łukaszowi Kaczyńskiemu.

Kraków, 18.12.2019 r.

Prof. dr hab. med. Anetta Undas