

Piotr Potemski<sup>1</sup>, Joanna Połowinczak-Przybyłek<sup>1</sup>, Rafał Wójcik<sup>2</sup>, Marcin Kaczor<sup>3</sup>

<sup>1</sup>Klinika Chemioterapii Nowotworów Uniwersytetu Medycznego w Łodzi, WWCOiT im. M. Kopernika w Łodzi

<sup>2</sup>Aestimo s.c., Kraków

<sup>3</sup>Uniwersytet Jagielloński Collegium Medicum w Krakowie

# Krytyczna ocena badań klinicznych w onkologii — część II

Critical appraisal of clinical trials in oncology — part II

**Artykuł jest tłumaczeniem pracy:**

Potemski P, Połowinczak-Przybyłek J, Wójcik R, Kaczor M. Critical appraisal of clinical trials in oncology — part II. *Oncol Clin Pract* 2019; 15: 158–166. DOI: 10.5603/OCP.2019.0013.

Należy cytować wersję pierwotną.

**Adres do korespondencji:**

Prof. dr hab. n. med. Piotr Potemski  
 Klinika Chemioterapii Nowotworów  
 Uniwersytet Medyczny w Łodzi;  
 WWCOiT im. M. Kopernika w Łodzi  
 e-mail: piotrho@mp.pl

**STRESZCZENIE**

Niniejsza praca stanowi drugą część cyklu przedstawiającego informacje pomocne w samodzielnej analizie wartości publikowanych wyników badań klinicznych w onkologii. W odniesieniu do wybranych badań klinicznych w onkologii podano w niej praktyczne przykłady krytycznej oceny ich założeń, konstrukcji, a także interpretacji wyników. Omówiono kilka badań typu *non-inferiority*. Zamieszczono przykłady publikacji, w których dokonywano analiz *post hoc*, grupowania zmiennych i wielokrotnych porównań. Przedstawiono także przykłady badań z kontrowersyjnym doбором chorych i komparatora, a także budzących wątpliwości co do tzw. istotności klinicznej uzyskanych wyników. Celem pracy było zwrócenie uwagi czytelnika na wybrane istotne elementy badań klinicznych i sposobu przedstawiania ich wyników, po to aby ułatwić w przyszłości lekarzom praktykom samodzielną ocenę dostępnych publikacji i przez to racjonalne wykorzystanie wyników badań klinicznych w codziennej praktyce.

**Słowa kluczowe:** onkologia, badania kliniczne, krytyczna ocena, analiza publikacji, metodologia badań, interpretacja wyników

**ABSTRACT**

The paper is the second part of the series presenting information useful for independent analysis of the value of published results of clinical trials in oncology. Based on selected examples of clinical trials, a few attempts of critical appraisal of clinical trials assumptions, construction and interpretation of their results were given. Several non-inferiority trials were discussed. The paper provides examples of publications in which post hoc analyses, grouping of variables and multiple comparisons were made. Examples of research with a controversial selection of patients and a comparator, as well as studies whose clinical significance of the obtained results is questionable are presented. The aim of our work was to draw the Reader's attention to selected essential elements of clinical trials and the way of presenting their results in order to facilitate practitioners the independent evaluation of available publications and rational use of clinical trial results in everyday practice in the future.

**Key words:** oncology, clinical trials, critical appraisal, publication analysis, research methodology, interpretation of results

**Wstęp**

W pierwszej części cyklu przedstawiono ogólne informacje pomocne w samodzielnej analizie wartości publikowanych wyników badań klinicznych w onkologii. Niestety, często opis metodyki badania jest przedstawia-

ny w publikacjach w sposób bardzo skrótowy, a więcej szczegółów można odnaleźć dopiero po zapoznaniu się z jego protokołem, który nie zawsze jest dostępny. Przy tym czas, jaki lekarz praktyk może poświęcić na krytyczną ocenę nowej publikacji, jest zwykle bardzo ograniczony. Wszystko to sprawia, że czytelnikowi, który nie

jest specjalistą w dziedzinie metodologii badań klinicznych, a lekarzem praktykiem, nawet po bardzo uważnej lekturze publikacji będącej efektem przeprowadzonej próby klinicznej, dość trudno jest systematycznie ocenić wszystkie elementy składające się na wiarygodność danego badania. W niniejszej pracy podano praktyczne przykłady interpretacji wybranych badań klinicznych. Z oczywistych względów przedstawione analizy nie mogą być uznane za kompleksową ocenę wartości wyników tych badań, a są jedynie próbą zwrócenia uwagi czytelnika na wybrane, ale w opinii autorów niniejszego opracowania istotne elementy, mogące wpłynąć na interpretację znaczenia opublikowanych wyników i ich roli w praktyce klinicznej.

### Badania *non-inferiority*

Badanie typu *non-inferiority*, ze względu na metodykę odmienną niż stosowana w powszechnie spotykanym badaniu typu *superiority*, z reguły przysparza klinicystom wielu problemów. Ocenia ono, czy dana interwencja nie jest gorsza w zakresie skuteczności klinicznej od obecnie stosowanego standardu leczenia. Podstawowym elementem podlegającym krytycznej ocenie podczas interpretacji tego rodzaju badań jest założona wartość delta, określająca możliwą do zaakceptowania różnicę skuteczności klinicznej porównywanych interwencji. Może być ona zdefiniowana poprzez np. określenie wielkości zachowania skuteczności klinicznej obecnego standardu leczenia na podstawie wyników historycznego badania porównującego obecny standard z leczeniem objawowym (badanie ASPECCT) lub wyznaczenie górnej granicy przedziału ufności na bazie przyjętej w ramach konsensusu wartości akceptowalnej klinicznie różnicy efektów (np. badanie IDEA). Interpretując wyniki tego rodzaju badań, warto zwrócić uwagę na to, jak duże mogą być założone różnice, uznawane wciąż za akceptowalne.

#### Badanie ASPECCT

Badanie ASPECCT było prospektywnym badaniem klinicznym III fazy typu *non-inferiority*, w którym planowano udowodnić, że monoterapia panitumumabem u wcześniej poddawanych chemioterapii chorych na uogólnionego raka jelita grubego pozwala uzyskać co najmniej połowę skuteczności cetuksymabu wyrażonej poprzez wydłużenie czasu całkowitego przeżycia (OS, *overall survival*) względem najlepszej terapii objawowej, wykazanego w historycznym badaniu III fazy [1]. Takie zdefiniowanie wartości delta świadczącej o *non-inferiority* wydaje się bardzo bezpiecznym założeniem, łatwym do potwierdzenia w próbie klinicznej. Badaniem, do którego odniesiono się, przyjmując to założenie, było badanie CO.17, którego wyniki w populacji pacjentów

bez mutacji *KRAS* opublikowano w roku 2008 [2]. W badaniu tym hazard względny (HR, *hazard ratio*) zgonu osób leczonych cetuksymabem w porównaniu z leczeniem wyłącznie objawowym wyniósł 0,55 [95% przedział ufności (CI, *confidence interval*): 0,41–0,74], a mediany OS, odpowiednio, 9,5 i 4,8 miesiąca. Do badania ASPECCT włączono 1010 chorych i już w pierwszej zaplanowanej analizie udowodniono przyjęte założenie, wykazując, że panitumumab zachowywał od 81,9 do 129,5% wpływu cetuksymabu na OS.

#### Badanie IDEA

Badanie IDEA (*The International Duration Evaluation of Adjuvant Therapy*) to prospektywnie zaplanowana łączna analiza indywidualnych danych pacjentów z sześciu badań z randomizacją, w których u chorych na raka okrężnicy porównywano skuteczność 3-miesięcznej chemioterapii adiuwantowej z oksaliplatyną (FOLFOX-4 albo zmodyfikowany FOLFOX-6, albo CAPOX) oraz standardowego leczenia trwającego 6 miesięcy [3]. Powodem zaplanowania takiego badania była chęć zmniejszenia toksyczności leczenia uzupełniającego (głównie polineuropatii) mogącej niekorzystnie wpływać na aktywność i jakość życia radykalnie leczonych pacjentów.

Pierwszorzędownym punktem końcowym był czas przeżycia wolnego od choroby (DFS, *disease-free survival*) w zmodyfikowanej populacji *intent-to-treat* (ITT; pacjenci poddani randomizacji, którzy otrzymali przynajmniej jedno podanie chemioterapii) — dotyczyło to 12 834 chorych spośród 13 025 poddanych randomizacji. Wartość delta ustalono jako górną granicę 95% CI HR<sub>DFS</sub>, wynoszącą 1,12. Jeśli zatem wartość 95% CI HR<sub>DFS</sub> przekroczyłaby 1,12, hipoteza zerowa nie mogłaby zostać odrzucona, co oznaczałoby, że leczenie krócej trwające jest gorsze o założoną wartość od leczenia standardowego. Taka wartość delta według autorów badania szacunkowo przekładała się na przewidywane zmniejszenie odsetka DFS po 3 latach o co najwyżej 2,7 punktu procentowego, a taka wartość została uznana przez badaczy jako możliwa do zaakceptowania. Dla przypomnienia, w innym badaniu dotyczącym raka okrężnicy (MOSAIC), w którym ustalono wartość chemioterapii uzupełniającej FOLFOX, pacjenci z III stopniem zaawansowania cechowali się odsetkiem DFS po 3 latach wynoszącym 72,2%, w porównaniu z 65,3% u otrzymujących fluorouracyl z folinianem wapniowym [4]. Istotniejszy wydaje się jednak wpływ FOLFOX wyrażony HR<sub>DFS</sub>, który w badaniu MOSAIC wyniósł 0,76 (95% CI: 0,62–0,92), co oznacza, że względny hazard nawrotu zmniejszył się o 24%, a „prawdziwa” wartość tego zmniejszenia (czyli wartość przenoszona na tzw. populację ogólną) zawierała się pomiędzy 8% a 38%. Przyjęta w badaniu IDEA wartość delta odpowiada zachowaniu ok. 60% efektu

w  $HR_{DFS}$  stwierdzonego dla porównania FOLFOX do fluorouracylu z folinianem wapniowym w grupie chorych z III stopniem zaawansowania uczestniczących w badaniu MOSAIC. W badaniu IDEA, zgodnie ze schematem randomizacji w wykorzystanych badaniach pierwotnych, porównywano interwencje wyłącznie pod względem czasu trwania chemioterapii, ale nie jej rodzaju.

W zmodyfikowanej populacji ITT  $HR_{DFS}$  pacjentów leczonych przez 3 miesiące w porównaniu z osobami leczonymi 6 miesięcy wyniósł 1,07 (95% CI: 1,0–1,15), założona wartość poziomu istotności statystycznej  $p$  dla hipotezy *non-inferiority* leczenia trwającego 3 miesiące wyniosła 0,11, a wartość  $p$  dla hipotezy *superiority* leczenia trwającego 6 miesięcy — 0,045. Oznacza to, że nie osiągnięto pierwszorzędowego punktu końcowego i nie udowodniono przy przyjętej wartości delta, że krótsze leczenie nie jest gorsze od standardowego. Niekiedy można napotkać ocenę wyników tego badania opartą na liczbowych wartościach odsetków DFS po 3 latach — wyniosły one, odpowiednio, 74,6% oraz 75,5%. Zgodnie z tą oceną różnica w odsetkach DFS (0,9 punktu procentowego) jest zbyt mała, aby miało to znaczenie kliniczne. Taka interpretacja wyników badania IDEA świadczy o pełnym niezrozumieniu metodyki statystycznej i jest całkowicie nieprawidłowa.

Warto zaznaczyć, że wcześniej zaplanowane analizy podgrup miały charakter eksploracyjny (czyli służyły generowaniu hipotez badawczych) i nie mogą być interpretowane w oderwaniu od zasadniczego wyniku badania. Stwierdzono, że na skuteczność leczenia trwającego 3 miesiące wpływa najprawdopodobniej rodzaj chemioterapii (FOLFOX albo CAPOX) ( $p$  testu interakcji miało wartość 0,006, a po wprowadzeniu korekty związanej z wielokrotnym testowaniem — 0,02). W grupie 5071 pacjentów otrzymujących CAPOX  $HR_{DFS}$  wyniósł 0,95 (95% CI: 0,85–1,06), co pozwoliłoby na potwierdzenie założeń *non-inferiority*. Niestety — jak wcześniej wspomniano — we włączonych badaniach nie dokonywano randomizacji w zależności od rodzaju chemioterapii ani nawet nie stratyfikowano randomizacji ze względu na rodzaj chemioterapii. Czynniki te sprawiają, że wynik dotyczący tylko schematu CAPOX może być zupełnie przypadkowy, zwłaszcza że trudno znaleźć uzasadnienie medyczne takiej obserwacji.

### Niezaplanowane wcześniej analizy podgrup i wielokrotne porównania

Jak przedstawiono w części pierwszej publikacji, randomizacja jest bardzo istotnym elementem prawidłowo zaprojektowanego i przeprowadzonego badania klinicznego. Umożliwia ona przy odpowiednio licznej populacji zapewnienie równomiernego rozkładu różnych — także nieznanych — czynników zakłócających.

Brak randomizacji albo częściowa utrata jej efektu, np. w wyniku przeprowadzania analiz *post hoc* dotyczących niezaplanowanych wcześniej podgrup pacjentów, oznacza, że porównywane grupy mogą znacząco się różnić rozkładem występowania innych istotnych cech rokowniczych.

#### Badanie IDEA

W wyżej omówionym badaniu IDEA wyjściowo zdefiniowano podgrupy w zależności od cechy T (T1–3 i T4) oraz N (N1 i N2) i w żadnej z nich nie wykazano założeń *non-inferiority* leczenia trwającego 3 miesiące. Analizując jednak wyniki badania *post hoc*, utworzono dwie kategorie ryzyka nawrotu — ryzyko małe (T1–3N1) oraz duże (T4 lub N2). W kategorii małego ryzyka (7471 chorych) granicznie potwierdzono założenie *non-inferiority* niezależnie od rodzaju chemioterapii ( $HR_{DFS} = 1,01$ ; 95% CI: 0,90–1,12); podobnie potwierdzono je w grupie chorych ( $n = 2852$ ) otrzymujących CAPOX i zaliczonych do kategorii małego ryzyka ( $HR_{DFS} = 0,85$ ; 95% CI: 0,71–1,01). We wszystkich pozostałych grupach chorych, tj. dużego ryzyka niezależnie od rodzaju chemioterapii lub małego ryzyka leczonych FOLFOX, nie udało się wykazać założeń *non-inferiority*. Należy wziąć pod uwagę, że oceny w zależności od tych kategorii nie zaplanowano wcześniej, a dokonano jej dopiero po analizie uzyskanych wyników. Oznacza to, że w podgrupach tych istotną rolę mogą odgrywać nieznane dodatkowe czynniki, które sprawiają, że wyników analiz *post hoc* nie można uznać za formalny dowód służący do wnioskowania o rzeczywistych różnicach w skuteczności interwencji. W opinii autorów niniejszej pracy jedyną potencjalnie użyteczną praktycznie sugestią wynikającą z tych analiz może być możliwość skrócenia do 3 miesięcy czasu trwania chemioterapii CAPOX u pacjentów z rakiem okrężnicy T1–3N1 w przypadku złej tolerancji leczenia jako alternatywa dla zmniejszenia dawki oksaliplatyny lub jej odstawienia i kontynuowania terapii samą fluoropirymidyną.

#### Badanie ASPECCT

Problem wielokrotnych porównań i przypadkowych wyników uznawanych za „istotne statystycznie” jeszcze lepiej ilustruje artykuł opublikowany w 2016 roku, w którym przedstawiono zaktualizowane wyniki badania ASPECCT oraz m.in. analizę *post hoc* w zależności od wcześniejszego leczenia bewacyzumabem [5]. Stwierdzono, że w grupie 258 pacjentów wcześniej leczonych bewacyzumabem OS był dłuższy, gdy otrzymywali panitumumab, a nie cetuksymab. Mediany OS wyniosły, odpowiednio, 11,3 i 9,8 miesiąca ( $HR = 0,75$ ; 95% CI: 0,58–0,97). Obserwacja ta przyczyniła się do podejmowania prób promowania panitumumabu, a nie cetuksymabu, jako leku z wyboru u chorych wcześniej

otrzymujących bewacyzumab. Nasuwa się pytanie, w jaki sposób można wyjaśnić przewagę panitumumabu nad cetuksymabem akurat u osób leczonych wcześniej bewacyzumabem, skoro mechanizm biologiczny działania obydwu leków jest bardzo podobny? Stanowi to dobry przykład niewłaściwej interpretacji wyników obserwacji, której natura jest najprawdopodobniej przypadkowa i nie powinna stanowić podstawy zmiany praktyki klinicznej. Oczywiście w takich sytuacjach granica istotności statystycznej powinna być mniejsza od zwykle przyjętej ( $< 0,05$ ), ponieważ wartość  $p$  musi uwzględniać napotykaną w analizach *post hoc*, niezaplanowane wielokrotne testowanie hipotezy (poprawka Bonferroniego).

Przedstawianie wyników wcześniej niezaplanowanych porównań oznacza, że najprawdopodobniej przeprowadzono wiele im podobnych, w innych podgrupach, i wybrano tylko niektóre z nich, ponieważ im więcej przeprowadzi się analiz *post hoc*, tym większe jest prawdopodobieństwo, że wynik któregoś z nich będzie całkowicie losowo „istotny statystycznie”.

Badania „wykazujące” znaczenie znaków zodiaku astrologicznego w medycynie

Bardzo pouczającymi przykładami pozornego wykazania nieistniejących w rzeczywistości związków są dwie prace opublikowane przez ich autorów właśnie po to, aby unaocznić czytelnikom niebezpieczeństwa wynikające z dokonywania wielokrotnych porównań i grupowania *post hoc* zmiennych [6, 7].

W pierwszej z nich oceniono związki między znakami zodiaku a 223 najczęstszymi przyczynami hospitalizacji mieszkańców Kanady [6]. Losowo podzielono grupę ponad 10 mln osób na dwie podgrupy — kohortę, w której testowano możliwe związki, i niezależną kohortę walidacyjną. Znalezione dwa znaki zodiaku wiążące się z większym ryzykiem hospitalizacji w porównaniu z pozostałymi dziesięcioma. W kohorcie walidacyjnej zbadano zależności między tymi dwoma znakami a poszczególnymi przyczynami hospitalizacji i stwierdzono, że osoby urodzone pod znakiem Lwa istotnie częściej ( $p = 0,0447$ ) były hospitalizowane z powodu krwawienia do przewodu pokarmowego, a urodzone pod znakiem Strzelca — istotnie częściej ( $p = 0,0123$ ) z powodu złamania kości ramiennej w porównaniu z osobami urodzonymi pod pozostałymi znakami zodiaku. Oczywiście po wprowadzeniu poprawki na wielokrotne testowanie te pozorne związki zniknęły.

W innej pracy badano zależność pomiędzy znakiem zodiaku a rokowaniem po chemioterapii mieloablacyjnej i przeszczepieniu allogenicznym komórek macierzystych układu krwiotwórczego u chorych na przewlekłą białaczkę szpikową [7]. Przeanalizowano prawdopodobieństwo przeżycia co najmniej 5 lat w grupie 626 pacjentów w zależności od ich znaku zodiaku i stwierdzono różnice

liczbowe, ale nieosiągające istotności statystycznej. Kiedy jednak wyodrębniono osoby urodzone pod znakiem Barana, Byka, Bliźniąt, Lwa, Skorpiona lub Koziorożca (łącznie 317 pacjentów) i porównano je z pozostałą grupą (309 osób), różnica okazała się znamieną statystycznie (5-letnie przeżycie 58% vs. 48%;  $p = 0,007$ ). Co więcej, po przeprowadzeniu analizy wieloczynnikowej uwzględniającej możliwy wpływ innych znanych czynników rokowniczych wyniki leczenia pacjentów urodzonych pod którymś z wymienionych znaków zodiaku nadal były istotnie lepsze niż u pozostałych osób ( $p = 0,005$ ). Autorzy opisywanej pracy podsumowują, że stanowi ona przykład „udowodnienia” nieistniejącego związku, a obserwowane „istotne” zależności są wynikiem grupowania zmiennych *post hoc* w celu uzyskania jak największej i „istotnej statystycznie” różnicy.

Badanie COU-AA-302

Przykładem właściwej interpretacji możliwego wpływu wielokrotnego testowania w odniesieniu do wyników analizy przeżycia jest przyjęcie innego progu istotności statystycznej dla wyników zaplanowanych wcześniej, etapowych analiz badań klinicznych z randomizacją. Na przykład, w badaniu III fazy COU-AA-302, w którym porównano abirateron w skojarzeniu z prednizonem do placebo z prednizonem u chorych na przerzutowego raka gruczołu krokowego opornego na kastrację, nieleczonych docetakselem, dwoma pierwszorzędownymi punktami końcowymi były: czas wolny od progresji radiologicznej (rPFS, *radiographic progression-free survival*) i OS. Typową wartość  $p$  równą 0,05 standardowo podzielono zatem na obydwa punkty końcowe — stwierdzenie istotności statystycznej różnicy w rPFS wymagało, aby  $p$  było mniejsze od 0,01, a w przypadku OS — mniejsze od 0,04 [8].

Zaplanowano przeprowadzenie oceny OS w kilku etapach (analizy *interim*) — po wystąpieniu, odpowiednio, 15%, 40% i 55% z wymaganej do ostatecznej analizy liczby zgonów — oraz analizy końcowej — po wystąpieniu co najmniej 773 zgonów (do badania włączono 1088 pacjentów). Ze względu na wielokrotne testowanie wpływu leku na OS (przy odcięciu danych w różnych punktach czasowych) zastosowano korektę granicznych wartości  $p$  wymaganych do stwierdzenia istotności statystycznej, odnotowywanych w tych etapowych analizach różnic zgodnie z procedurą opisaną przez O’Briana i Fleminga. Pierwsza publikacja zawierała ostateczny wynik analizy rPFS, w której stwierdzono istotną różnicę między abirateronem a placebo ( $HR = 0,55$ ; 95% CI: 0,45–0,62;  $p < 0,001$ ), i wynik etapowej analizy OS po wystąpieniu 43% z wymaganych 773 zdarzeń. Stwierdzono, że  $HR_{OS}$  wynosił 0,75 (95% CI: 0,61–0,93;  $p = 0,01$ ). Mimo że wartość  $p$  była mniejsza niż wymagane 0,04, to do stwierdzenia istotności sta-

tystycznej różnicy w OS w tej analizie etapowej upoważniała dopiero skorygowana wartość  $p$  wynosząca 0,001 lub mniej. Wynik kolejnej opublikowanej analizy etapowej przeprowadzonej po stwierdzeniu 56% zgonów także nie był istotny statystycznie (HR = 0,79; 95% CI: 0,66–0,95;  $p = 0,0151$ ; wymagana wartość skorygowana  $p = 0,0035$ ) [9]. Dopiero ostateczna analiza OS wykonana po stwierdzeniu 96% z 773 zgonów umożliwiła stwierdzenie wpływu leku na wydłużenie OS (HR 0,81; 95% CI: 0,70–0,93;  $p = 0,0033$ ), czyli na osiągnięcie drugiego pierwszorzędowego punktu końcowego [10].

## Dobór pacjentów i komparatora

Prawidłowy dobór chorych, odpowiedni do przeprowadzenia planowanej interwencji, jest niezbędnym elementem dobrze zaplanowanego i wykonanego badania, a także pozwala na ekstrapolację jego wyników na populację, która będzie leczona w warunkach rzeczywistej praktyki klinicznej. Zastosowanie właściwego komparatora, będące kluczowym elementem prawidłowo przeprowadzonego badania klinicznego, zakłada wykorzystanie terapii, która jest zgodna z aktualną praktyką kliniczną oraz powszechnie przyjętymi zaleceniami i wytycznymi, z uwzględnieniem ich ciągłej ewolucji, szczególnie w dziedzinie tak szybko się rozwijającej jak onkologia. Dobór nieodpowiedni, sprzeczny z powszechnie przyjętymi zaleceniami uniemożliwia przyjęcie wniosków z badania jako wiarygodne (wiarygodność zewnętrzna). Przykładem niedawno opublikowanego badania z bardzo kontrowersyjnym doбором chorych jest CARMENA.

### Badanie CARMENA

Celem opublikowanego w roku 2018 badania III fazy typu *non-inferiority* CARMENA (*Cancer du Rein Metastatique Nephrectomie et Antiangiogéniques*) było wykazanie, że niewykonywanie nefrektomii u pacjentów z przerzutowym rakiem nerki przed leczeniem sunitynibem nie jest gorsze od takiej terapii poprzedzonej nefrektomią [11]. Pierwszorzędowym punktem końcowym był OS, a randomizację stratyfikowano m.in. w zależności od kategorii rokowniczej. Wyniki badania CARMENA są dość powszechnie interpretowane jako mogące zmienić praktykę kliniczną, ponieważ wykazano, że niewykonywanie nefrektomii nie jest gorszym postępowaniem (HR<sub>OS</sub> = 0,89; 95% CI: 0,71–1,10; kryterium *non-inferiority*: górna granica 95% CI nie większa niż 1,20).

Analizując znaczenie uzyskanego wyniku i jego potencjalny wpływ na praktykę kliniczną, należy jednak pamiętać o podstawowym ograniczeniu tego badania, jakim był dobór pacjentów. Do badania włączano

bowiem chorych spełniających kryteria pośredniego lub niekorzystnego rokowania według klasycznej skali MSKCC (*Memorial Sloan Kettering Cancer Center*) i aż 43% pacjentów uczestniczących w CARMENA było przypisanych do kategorii niekorzystnego rokowania. U takich chorych do niedawna jedynym lekiem, dla którego w badaniu III fazy wykazano niewielki wpływ na wydłużenie OS, był temsyrolimus — lek obecnie refundowany u pacjentów zaliczanych do takiej kategorii, także w Polsce [12]. Nie ma też żadnych wiarygodnych danych pochodzących z badania z randomizacją potwierdzających wpływ zastosowanego w badaniu CARMENA sunitynibu na OS u pacjentów z niekorzystnym rokowaniem. Ponadto, u takich chorych nefrektomia z reguły nie jest wykonywana i zalecana, a we wspomnianym badaniu III fazy z wykorzystaniem temsyrolimusu nie odnotowano jakiegokolwiek pozytywnego wpływu tej procedury (wykonywanej niekiedy na długo przed randomizacją) na skuteczność inhibitora mTOR. Z tych powodów umożliwienie włączania do badania oceniającego wpływ zaniechania wykonywania nefrektomii pacjentów z grupy niekorzystnego rokowania i leczenie ich sunitynibem należy uznać za postępowanie nieuzasadnione istniejącą wiedzą medyczną. Stosowanie w niektórych ośrodkach sunitynibu, a nie temsyrolimusu u chorych z niekorzystnym rokowaniem nie może być uznane za praktykę opartą na *evidence based medicine*.

Wyniki badania CARMENA zostały uzyskane u wszystkich włączonych do niego pacjentów i osoby z niekorzystnym rokowaniem miały istotny wpływ na jego końcowy rezultat. W tej właśnie grupie ani nefrektomia, ani sunitynib mogły w ogóle nie wpływać na pierwszorzędowy punkt końcowy. Przy takim założeniu udowodnienie u pacjentów z niekorzystnym rokowaniem *non-inferiority* niewykonywania nefrektomii nie byłoby trudne. Wskazówką zdającą się potwierdzać taką hipotezę mogą być wyniki dotyczące każdej z grup rokowniczych osobno. W grupie niekorzystnego rokowania HR<sub>OS</sub> wyniósł 0,86 (95% CI: 0,62–1,17), a pośredniego — 0,92 (95% CI: 0,68–1,24). Jak widać, tylko w grupie niekorzystnego rokowania uzyskany wynik spełnił przyjęte kryterium *non-inferiority* (górna granica 95% CI nie większa niż 1,20). Oczywiście nie może to być uznane za dowód, a jedynie za przesłankę wskazującą na poprawność podanej interpretacji.

Analiza statystyczna przeprowadzona w populacji ITT zakłada ocenę wyników wszystkich chorych poddanych randomizacji, bez względu na to, czy otrzymali przypisaną interwencję czy nie. Interpretacja wyniku badania CARMENA powinna także uwzględniać fakt, że u 16 pacjentów z grupy 226 osób włączonych do ramienia z nefrektomią (7%) nie wykonano jej, a 38 z 224 osób poddanych randomizacji do ramienia tylko z terapią systemową (17%) zostało poddanych nefrektomii. Powoduje to zmniejszenie różnic między

ramionami i ułatwia wykazanie *non-inferiority* w populacji ITT. Wreszcie, pierwotny plan zakładał włączenie 576 chorych i ocenę po wystąpieniu 452 zgonów, jednak w wyniku niesatysfakcjonującego tempa naboru badanie przerwano po włączeniu do niego 450 osób, a ostateczne wyniki opublikowano po wystąpieniu 326 zgonów.

#### Badanie CheckMate 214

Problem doboru komparatora i populacji docelowej napotykaemy także w badaniu III fazy CheckMate 214, w którym oceniono skojarzenie niwolumabu z ipilimumabem u chorych na przerzutowego raka nerki i jako komparator zastosowano sunitynib [13]. O ile u chorych z kategorią korzystnego lub pośredniego rokowania taki komparator nie budzi wątpliwości, o tyle u pacjentów zaliczonych do kategorii niekorzystnego rokowania trudno go — z powodów wcześniej omówionych — uznać za optymalny. Tacy chorzy stanowili aż 21% populacji, w której oceniano pierwszorzędowe punkty końcowe, tj. odsetek obiektywnych odpowiedzi, czas wolny od progresji (PFS, *progression free survival*) i OS. Wartość poziomu istotności statystycznej  $p < 0,05$  podzielono na: 0,001 (odpowiedzi obiektywne), 0,009 (PFS) i 0,04 (OS). Do badania włączano chorych niezależnie od kategorii rokowniczej, ale ocenę pierwszorzędowych punktów końcowych zaplanowano wyłącznie u chorych zaliczonych do kategorii pośredniego bądź niekorzystnego rokowania. Drugorzędowymi punktami końcowymi były: odsetek obiektywnych odpowiedzi, PFS i OS u osób poddanych randomizacji (populacja ITT) oraz częstość zdarzeń niepożądanych u pacjentów, którzy zostali poddani leczeniu. Zaplanowano jedynie dokonanie eksploratywnej analizy w grupie 249 chorych z kategorią korzystnego rokowania (21% populacji osób poddanych randomizacji).

W odniesieniu do pierwszorzędowych punktów końcowych uzyskano różnice w odsetku odpowiedzi obiektywnych i OS na korzyść immunoterapii, ale nie w przypadku PFS ( $p = 0,03$ ). W populacji osób poddanych randomizacji nie uzyskano istotnej różnicy w żadnym z trzech drugorzędowych punktów końcowych oceniających skuteczność. Oznacza to, że dołączenie chorych z grupy korzystnego rokowania wniosło korzystny efekt immunoterapii w odniesieniu do odsetka odpowiedzi i OS. Bardzo niepokojące są wyniki eksploratywnej analizy dotyczącej skuteczności w grupie korzystnego rokowania — nastąpiło tu odwrócenie wpływu immunoterapii i komparatora na niekorzyść leczenia eksperymentalnego. Odsetek odpowiedzi wyniósł 29% vs. 52% ( $p < 0,001$ ),  $HR_{PFS} = 2,18$  ( $p < 0,001$ ), a  $HR_{OS} = 1,45$  ( $p = 0,27$ , przy czym wystąpiło tylko 37 zgonów).

Oczywista interpretacja wyników badania wskazuje, że korzyść z immunoterapii ograniczona jest wyłącznie do pacjentów z kategorią pośredniego lub niekorzystnego

rokowania (z wyraźnym zastrzeżeniem odnośnie do wiarygodności zewnętrznej w związku z zastosowaniem nieoptymalnego komparatora w tej drugiej grupie). Należy jednak zauważyć, że z korzyścią z immunoterapii wiązało się wystąpienie tylko jednego z sześciu czynników niekorzystnego rokowania według klasyfikacji IMDC (*International Metastatic Renal Carcinoma Database Consortium*), takich jak stan sprawności według Karnofsky'ego 70, czas od rozpoznania nowotworu do randomizacji krótszy niż rok, niedokrwistość, skorygowane stężenie wapnia w surowicy powyżej 10 mg/dl, granulocytoza obojętnochłonna i nadpłytkowość. Nieuchronnie rodzą się zatem uzasadnione wątpliwości, czy związek ten jest prawdziwy dla każdego z wymienionych — tak różnych — czynników i czy korzyść z immunoterapii zależy od liczby czynników niekorzystnego rokowania? Niestety publikacja wyników badania CheckMate 214 nie pozwala odpowiedzieć na te pytania. Wśród 667 pacjentów należących do grupy pośredniego ryzyka nie przeprowadzono analiz mogących wyjaśnić wspomniane wątpliwości.

Kolejnym zaskakującym wyborem było zastosowanie w ramieniu eksperymentalnym wyłącznie skojarzenia przeciwciała anti-CTLA4 i anti-PD1, ale nie monoterapii anti-PD1. Mogło być to spowodowane chęcią uzyskania jak najlepszego efektu bezpośredniego (odsetek odpowiedzi obiektywnych był jednym z pierwszorzędowych punktów końcowych). Nie podważa to wartości samego skojarzenia, rodzi jednak pytanie, czy monoterapia anti-PD1 nie byłaby równie skuteczna, a przy tym mniej toksyczna? Pytanie to można uznać za bardzo zasadne zwłaszcza w kontekście opublikowanych pod koniec 2015 roku ostatecznych wyników wcześniej rozpoczętego badania CheckMate 025, w którym u wcześniej leczonych pacjentów zastosowano z dobrym wynikiem właśnie monoterapię niwolumabem. Taka wątpliwość była wskazywana przez CHMP (*Committee for Medicinal Products for Human Use*) Europejskiej Agencji Leków (EMA, *European Medicines Agency*) jako uzasadnienie zaskakującej pierwotnej negatywnej opinii dotyczącej rejestracji tego skojarzenia u chorych na raka nerki.

#### Badanie SOLO3

Zdumiewający jest również dobór komparatora w trwającym badaniu III fazy SOLO3, w którym u kobiet z nawrotowym (przynajmniej 2 wcześniejsze linie chemioterapii zawierające związek platyny), jednak nadal wrażliwym na związku platyny (progresja następująca później niż po 6 miesiącach od zakończenia ostatniej chemioterapii) rakiem jajnika z obecnością mutacji germlinalnej w *BRCA* olaparab porównywany jest do monochemioterapii [14]. Zaplanowany pierwszorzędowy punkt końcowy badania stanowi odsetek obiektywnych odpowiedzi. Komparator to wybór badacza pomiędzy

paklitaksellem, liposomalną doksorubicyną, topotekaniem i gemcytabiną.

Dla przypomnienia, powszechnie obowiązującym standardem leczenia takich chorych jest chemioterapia zawierająca związek platyny, a nie monoterapia którymś z wymienionych leków. Poza tym, wybór pierwszorzędnego punktu końcowego także trudno uznać za właściwy i klinicznie ważny. Biorąc pod uwagę te zastrzeżenia, trudno sobie wyobrazić, aby wynik tego badania mógł być użyteczny w praktyce klinicznej.

#### Badanie 20100007

O ile zastosowanie placebo lub wyłącznie najlepszego leczenia objawowego jako komparatora jest powszechną praktyką w przypadku, gdy nie występuje inna poza leczeniem objawowym opcja terapeutyczna o udowodnionej wartości (zwykle ostatnia linia leczenia) lub badana interwencja i placebo jako komparator są dodane do aktualnego standardu (*add-on*), o tyle zastosowanie placebo lub tylko najlepszego leczenia objawowego w sytuacji, gdy istnieją inne terapie o wcześniej wykazanej skuteczności, zawsze musi budzić wątpliwości natury etycznej. Celem tego opublikowanego w 2016 roku badania III fazy było wykazanie, że panitumumab wpływa na wydłużenie OS w porównaniu z najlepszym leczeniem objawowym u wcześniej leczonych systemowo chorych z przerzutowym rakiem jelita grubego bez mutacji w eksonie 2. genu *KRAS* [15]. W projekcie tego badania nie byłoby niczego zaskakującego, gdyby nie fakt, że nabór pacjentów przeprowadzono między listopadem 2011 roku a lipcem 2013 roku, proponując im najlepsze leczenie objawowe jako komparator. Było to sprzeczne z powszechną już wiedzą o skuteczności cetuksymabu w leczeniu chorych z przerzutowym rakiem jelita grubego. W opublikowanym w 2007 roku badaniu III fazy, do którego nabór rozpoczął się w styczniu roku 2004 i w którym po raz pierwszy oceniono wartość panitumumabu, porównując to przeciwiało do najlepszego leczenia objawowego, autorzy podkreślili, że w projekcie od początku założono możliwość zmiany ramienia badania w grupie kontrolnej po stwierdzeniu progresji (*cross-over*) ze względu na „znaną wcześniej aktywność panitumumabu i cetuksymabu” [16]. Dodatkowo, projekt tego badania z PFS jako pierwszorzędnym punktem końcowym i założonym *cross-over* spowodował, że podczas rejestracji nałożono na producenta wymóg zaplanowania i przeprowadzenia badania *non-inferiority* porównującego panitumumab z cetuksymabem, ponieważ od 2007 roku znany był wpływ cetuksymabu na wydłużenie OS w porównaniu z najlepszym leczeniem objawowym. Obydwa przeciwiała, tj. panitumumab i cetuksymab, zostały zarejestrowane do leczenia chorych z chemioopornym przerzutowym rakiem jelita grubego zarówno przez amerykańską Agencję ds. Żywności

i Leków (FDA, *Food and Drug Administration*), jak i EMA. W Unii Europejskiej cetuksymab dopuszczono do obrotu w 2004 roku, a panitumumab — w 2007 roku (już z uwzględnieniem stanu *KRAS*). W roku 2008 rejestrację cetuksymabu zmodyfikowano, uwzględniając stan *KRAS*. W lutym 2010 roku rozpoczęto nabór pacjentów do omówionego wcześniej badania *non-inferiority* ASPECCT, w którym OS był pierwszorzędnym punktem końcowym. W jaki sposób zatem kilka lat po pierwszych rejestracjach obydwu leków możliwe było przeprowadzenie badania klinicznego, w którym połowa włączonych do badania pacjentów otrzymała tylko najlepsze leczenie objawowe, czyli postępowanie skrajnie suboptymalne? Oczywiście w badaniu 20100007 nie przewidziano *cross-over* do panitumumabu po stwierdzeniu progresji w grupie kontrolnej, a przecież taka możliwość była w badaniu, które prowadzono w okresie, kiedy jeszcze nie było danych potwierdzających wpływ leków anty-EGFR na OS.

#### Czy każda różnica istotna statystycznie jest także istotna klinicznie?

Elementem krytycznej oceny badań klinicznych w onkologii powinna być zawsze odpowiedź na pytanie, czy uzyskana w badaniu klinicznym istotność statystyczna ma znaczenie praktyczne i wystarczy do zmiany praktyki klinicznej. To zagadnienie budzi wiele kontrowersji, ponieważ ocena, jaki zysk z przedłużenia PFS lub OS można uznać za istotny klinicznie, jest skrajnie subiektywna. Dyskusja na ten temat rozpoczęła się między innymi dlatego, że pojawiła się tendencja do projektowania komercyjnych badań obejmujących bardzo duże grupy chorych, w których można było wykazać bardzo niewielkie różnice w skuteczności, osiągające jednak poziom istotności statystycznej. Jeśli uwzględnić przy tym fakt, że pierwszorzędnym punktem końcowym tych badań nierzadko był PFS, trudno było przełożyć ich wyniki na praktykę kliniczną, zwłaszcza przy uwzględnieniu z reguły większej toksyczności i znacznego kosztu nowych leków.

#### Badania NCIC CTG PA.3 i VELOUR

W badaniu III fazy NCIC CTG PA.3 wykazano, że skojarzenie gemcytabiny i erlotynibu u chorych z zaawansowanym rakiem trzustki wpływa na istotne statystycznie wydłużenie OS [17]. Mimo że było to pierwsze badanie III fazy wskazujące na korzyść w OS z dołączenia do gemcytabiny innego leku, dość powszechny odbiór znaczenia wyniku tego badania dla praktyki klinicznej był daleki od entuzjazmu i badanie to stało się klasycznym przykładem wyniku istotnego statystycznie, ale nieistotnego klinicznie. Przyczyną tego

była przede wszystkim małą liczbowo różnicą w OS — HR miał wartość 0,82 (95% CI: 0,69–0,99), mediana OS 6,24 vs. 5,91 miesiąca, a prawdopodobieństwo przeżycia 12 miesięcy 23% vs. 17%. Duża liczba włączonych do badania chorych (569 osób) sprawiła, że bezwzględna różnica w liczbie zgonów między ramionami, wynosząca 8 przypadków, przełożyła się na istotność statystyczną testu log-rank. Szczególnie podkreślana przez komentatorów była różnica w medianach, wynosząca tylko 0,33 miesiąca. Ponadto w ramieniu eksperymentalnym obserwowano większą częstość niektórych zdarzeń niepożądanych, np. biegunki i zmian skórnych.

Ocena wartości badania NCIC CTG PA.3 tylko przez przyrównanie różnic w medianach, mimo że łatwa do zakomunikowania, jest oczywiście dość uproszczona, ponieważ obejmuje tylko jeden punkt czasowy przebiegu krzywych przeżycia. Lepszą, choć nieintuicyjną miarę stanowi hazard względny zgonu, którego zmniejszenie o 18% jest już bardziej obiecujące klinicznie. Dla porównania, aflibercept, lek obecnie refundowany w Polsce w 2. linii leczenia chorych na przerzutowego raka jelita grubego, dodany do chemioterapii FOLFIRI w badaniu III fazy VELOUR wpłynął na zmniejszenie względnego hazardu zgonu także o 18% (HR 0,817, 95% CI: 0,713–0,937). Różnica w medianach wynosiła 1,44 miesiąca, a prawdopodobieństwo przeżycia 24-miesiący — 28% vs. 19% [18]. Wykazanie, że taka różnica jest istotna statystycznie na poziomie  $p = 0,0032$ , było możliwe dzięki włączeniu do badania aż 1226 pacjentów. Następnie opublikowano kilka doniesień mających rozwiązać wątpliwości, czy stwierdzona w badaniu VELOUR różnica w rokowaniu jest istotna klinicznie. Rok po oryginalnej publikacji, ekstrapolując rzeczywiście uzyskane dane poza czas trwania badania przy użyciu metod matematycznych, oszacowano w perspektywie 15 lat średnie czasy przeżycia pacjentów z obydwu ramion, stwierdzając, że różnica między nimi wynosi 4,7 miesiąca, co — jak się wydaje — miało wywrzeć lepsze wrażenie na ówczesnych czytelnikach niż różnica w medianach 1,44 miesiąca [19]. Kolejna próba, której wynik ogłoszono drukiem w 2014 roku, polegała na dokonaniu analiz *post hoc* i wyodrębnieniu w ich wyniku podgrup odnoszących „większą” korzyść z leczenia eksperymentalnego [20]. Stwierdzono, że pacjenci w bardzo dobrym stanie sprawności (PS 0) z dowolną liczbą miejsc przerzutów odległych oraz pacjenci w dobrym stanie sprawności (PS 1) z przerzutami w tylko jednej lokalizacji cechują się medianą OS większą o 3,1 miesiąca, jeśli otrzymywali aflibercept. Tego typu analizy mogą oczywiście generować hipotezy badawcze, ale z pewnością nie pozwalają na przeniesienie uzyskanych w ten sposób rezultatów na tzw. populację ogólną. Wartość analiz *post hoc*, polegających na grupowaniu zmiennych, została już omówiona w niniejszym artykule. W tym miejscu należy wspomnieć, że takie działania wykorzystywane są niestety do uzyska-

nia jak największej liczbowo różnicy w medianach czasu przeżycia, co ułatwia otrzymanie korzystniejszych wyników analizy efektywności kosztowej przeprowadzanej w ramach procesu wnioskowania o refundację. Wielce prawdopodobne jest to, że wyodrębnione w taki sposób podgrupy mogą nie mieć żadnej rzeczywistej wartości predykcyjnej.

#### Badanie NO16966

Przykładem badania, które przynajmniej w niektórych krajach (np. w USA) wpłynęło na zmianę praktyki klinicznej pomimo budzącej poważne wątpliwości rzeczywistej wartości uzyskanych wyników, było badanie III fazy oceniające korzyść z dodania bewacyzumabu do chemioterapii zawierającej oksaliplatynę w 1. linii leczenia chorych na przerzutowego raka jelita grubego [21]. Pierwszorzędnym punktem końcowym był PFS i pacjenci, którzy otrzymywali chemioterapię FOLFOX-4 albo XELOX z bewacyzumabem, cechowali się dłuższym PFS (HR = 0,83, 95% CI: 0,72–0,95, mediana PFS: 9,4 vs. 8,0 miesiąca). Różnica ta była istotna statystycznie na poziomie  $p = 0,0023$ , a jej wykazanie było możliwe ze względu na liczebność grupy badanej obejmującej 1401 chorych. Naturalnie nie stwierdzono wpływu przeciwnie na OS.

Wątpliwości co do oceny znaczenia klinicznego wyników niektórych badań sprawiły, że m.in. ESMO (*European Society for Medical Oncology*) zaproponowało mogące być uznane za istotne klinicznie wartości różnic w poszczególnych punktach końcowych w zależności m.in. od rokowania [22]. W przeglądzie opublikowanych w latach 2011–2015 badań klinicznych z randomizacją dotyczących systemowego leczenia pacjentów z rakiem piersi, niedrobnokomórkowym rakiem płuca, rakiem jelita grubego lub trzustki uwzględniono 277 badań [23]. W 138 spośród nich przedstawiono istotne statystycznie różnice między terapią eksperymentalną a komparatorem, jednak po zastosowaniu proponowanych przez ESMO kryteriów istotności klinicznej uznano za takie wyniki tylko 43 (31%) ze 138 badań.

#### Podsumowanie

W niniejszej pracy autorzy skoncentrowali się na wybranych zagadnieniach, ilustrując je przykładami konkretnych badań klinicznych. Omówiono badania typu *non-inferiority*, ponieważ ten rodzaj badań klinicznych z reguły przysparza czytelnikom wielu problemów, co jest związane z metodyką zupełnie odmienną w porównaniu z metodyką badań, których celem jest wykazanie wyższości jednej interwencji nad inną. Podano przykłady publikacji, w których dokonywano analiz *post hoc*, grupowania zmiennych i wielokrotnych porównań.



Przedstawiono przykłady badań klinicznych, których zrozumienie i interpretacja są niemożliwe bez zwrócenia uwagi na wątpliwości co do charakterystyki włączanych pacjentów albo doboru komparatora. Skrajny przykład badań, których wyniki trudno przenieść do praktyki klinicznej, stanowią takie, w których grupa kontrolna leczona jest nieoptymalnie, czyli mniej skutecznie, niż jest to możliwe. Takich badań na szczęście nie ma dużo, znacznie częściej natomiast zdarzają się próby kliniczne, w których wątpliwości dotyczą części włączonych do nich chorych. W opracowaniu podano również przykłady badań budzących wątpliwości co do tzw. istotności klinicznej ich wyników.

Autorzy mają nadzieję, że lektura cyklu dwóch prac będących efektem współpracy statystyków medycznych i lekarzy onkologów ułatwi czytelnikom samodzielną interpretację dostępnych publikacji i przez to racjonalne wykorzystanie wyników badań klinicznych w codziennej praktyce.

## Piśmiennictwo

- Price TJ, Peeters M, Kim TW, et al. Panitumumab versus cetuximab in patients with chemotherapy-refractory wild-type KRAS exon 2 metastatic colorectal cancer (ASPCCCT): a randomised, multicentre, open-label, non-inferiority phase 3 study. *Lancet Oncol.* 2014; 15(6): 569–579, doi: [10.1016/S1470-2045\(14\)70118-4](https://doi.org/10.1016/S1470-2045(14)70118-4), indexed in Pubmed: [24739896](https://pubmed.ncbi.nlm.nih.gov/24739896/).
- Karapetis CS, Khambata-Ford S, Jonker DJ, et al. K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *N Engl J Med.* 2008; 359(17): 1757–1765, doi: [10.1056/NEJMoa0804385](https://doi.org/10.1056/NEJMoa0804385), indexed in Pubmed: [18946061](https://pubmed.ncbi.nlm.nih.gov/18946061/).
- Grothey A, Sobrero AF, Shields AF, et al. Duration of adjuvant chemotherapy for stage III colon cancer. *N Engl J Med.* 2018; 378(13): 1177–1188, doi: [10.1056/NEJMoa1713709](https://doi.org/10.1056/NEJMoa1713709), indexed in Pubmed: [29590544](https://pubmed.ncbi.nlm.nih.gov/29590544/).
- André T, Boni C, Mounedji-Boudiaf L, et al. Oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment for colon cancer. *N Engl J Med.* 2004; 350(23): 2343–2351, doi: [10.1056/NEJMoa032709](https://doi.org/10.1056/NEJMoa032709), indexed in Pubmed: [15175436](https://pubmed.ncbi.nlm.nih.gov/15175436/).
- Price T, Kim TW, Li J, et al. Final results and outcomes by prior bevacizumab exposure, skin toxicity, and hypomagnesaemia from ASPCCCT: randomized phase 3 non-inferiority study of panitumumab versus cetuximab in chemorefractory wild-type KRAS exon 2 metastatic colorectal cancer. *Eur J Cancer.* 2016; 68: 51–59, doi: [10.1016/j.ejca.2016.08.010](https://doi.org/10.1016/j.ejca.2016.08.010), indexed in Pubmed: [27716478](https://pubmed.ncbi.nlm.nih.gov/27716478/).
- Austin PC, Mamdani MM, Juurlink DN, et al. Testing multiple statistical hypotheses resulted in spurious associations: a study of astrological signs and health. *J Clin Epidemiol.* 2006; 59(9): 964–969, doi: [10.1016/j.jclinepi.2006.01.012](https://doi.org/10.1016/j.jclinepi.2006.01.012), indexed in Pubmed: [16895820](https://pubmed.ncbi.nlm.nih.gov/16895820/).
- Szydło RM, Gabriel I, Olavarria E, et al. Sign of the Zodiac as a predictor of survival for recipients of an allogeneic stem cell transplant for chronic myeloid leukaemia (CML): an artificial association. *Transplant Proc.* 2010; 42(8): 3312–3315, doi: [10.1016/j.transproceed.2010.07.036](https://doi.org/10.1016/j.transproceed.2010.07.036), indexed in Pubmed: [20970679](https://pubmed.ncbi.nlm.nih.gov/20970679/).
- Ryan CJ, Smith MR, de Bono JS, et al. COU-AA-302 Investigators. Abiraterone in metastatic prostate cancer without previous chemotherapy. *N Engl J Med.* 2013; 368(2): 138–148, doi: [10.1056/NEJMoa1209096](https://doi.org/10.1056/NEJMoa1209096), indexed in Pubmed: [23228172](https://pubmed.ncbi.nlm.nih.gov/23228172/).
- Rathkopf DE, Smith MR, de Bono JS, et al. Updated interim efficacy analysis and long-term safety of abiraterone acetate in metastatic castration-resistant prostate cancer patients without prior chemotherapy (COU-AA-302). *Eur Urol.* 2014; 66(5): 815–825, doi: [10.1016/j.eururo.2014.02.056](https://doi.org/10.1016/j.eururo.2014.02.056), indexed in Pubmed: [24647231](https://pubmed.ncbi.nlm.nih.gov/24647231/).
- Ryan C, Smith M, Fizazi K, et al. Abiraterone acetate plus prednisone versus placebo plus prednisone in chemotherapy-naïve men with metastatic castration-resistant prostate cancer (COU-AA-302): final overall survival analysis of a randomised, double-blind, placebo-controlled phase 3 study. *Lancet Oncol.* 2015; 16(2): 152–160, doi: [10.1016/S1470-2045\(14\)71205-7](https://doi.org/10.1016/S1470-2045(14)71205-7).
- Méjean A, Ravaud A, Thezenas S, et al. Sunitinib Alone or after Nephrectomy in Metastatic Renal-Cell Carcinoma. *N Engl J Med.* 2018; 379(5): 417–427, doi: [10.1056/NEJMoa1803675](https://doi.org/10.1056/NEJMoa1803675), indexed in Pubmed: [29860937](https://pubmed.ncbi.nlm.nih.gov/29860937/).
- Hudes G, Carducci M, Tomczak P, et al. Global ARCC Trial. Temsirolimus, interferon alfa, or both for advanced renal-cell carcinoma. *N Engl J Med.* 2007; 356(22): 2271–2281, doi: [10.1056/NEJMoa066838](https://doi.org/10.1056/NEJMoa066838), indexed in Pubmed: [17538086](https://pubmed.ncbi.nlm.nih.gov/17538086/).
- Motzer RJ, Tannir NM, McDermott DF, et al. CheckMate 214 Investigators. Nivolumab plus ipilimumab versus sunitinib in advanced renal-cell carcinoma. *N Engl J Med.* 2018; 378(14): 1277–1290, doi: [10.1056/NEJMoa1712126](https://doi.org/10.1056/NEJMoa1712126), indexed in Pubmed: [29562145](https://pubmed.ncbi.nlm.nih.gov/29562145/).
- <https://clinicaltrials.gov/ct2/show/NCT02282020>.
- Kim TW, Elme A, Kucic Z, et al. A phase 3 trial evaluating panitumumab plus best supportive care vs best supportive care in chemorefractory wild-type KRAS or RAS metastatic colorectal cancer. *Br J Cancer.* 2016; 115(10): 1206–1214, doi: [10.1038/bjc.2016.309](https://doi.org/10.1038/bjc.2016.309), indexed in Pubmed: [27736842](https://pubmed.ncbi.nlm.nih.gov/27736842/).
- Van Cutsem E, Peeters M, Siena S, et al. Open-label phase III trial of panitumumab plus best supportive care compared with best supportive care alone in patients with chemotherapy-refractory metastatic colorectal cancer. *J Clin Oncol.* 2007; 25(13): 1658–1664, doi: [10.1200/JCO.2006.08.1620](https://doi.org/10.1200/JCO.2006.08.1620), indexed in Pubmed: [17470858](https://pubmed.ncbi.nlm.nih.gov/17470858/).
- Moore MJ, Goldstein D, Hamm J, et al. National Cancer Institute of Canada Clinical Trials Group. Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advanced pancreatic cancer: a phase III trial of the National Cancer Institute of Canada Clinical Trials Group. *J Clin Oncol.* 2007; 25(15): 1960–1966, doi: [10.1200/JCO.2006.07.9525](https://doi.org/10.1200/JCO.2006.07.9525), indexed in Pubmed: [17452677](https://pubmed.ncbi.nlm.nih.gov/17452677/).
- Van Cutsem E, Tabernero J, Lakomy R, et al. Addition of aflibercept to fluorouracil, leucovorin, and irinotecan improves survival in a phase III randomized trial in patients with metastatic colorectal cancer previously treated with an oxaliplatin-based regimen. *J Clin Oncol.* 2012; 30(28): 3499–3506, doi: [10.1200/JCO.2012.42.8201](https://doi.org/10.1200/JCO.2012.42.8201), indexed in Pubmed: [22949147](https://pubmed.ncbi.nlm.nih.gov/22949147/).
- Joulain F, Proskorovsky I, Allegra C, et al. Mean overall survival gain with aflibercept plus FOLFIRI vs placebo plus FOLFIRI in patients with previously treated metastatic colorectal cancer. *Br J Cancer.* 2013; 109(7): 1735–1743, doi: [10.1038/bjc.2013.523](https://doi.org/10.1038/bjc.2013.523), indexed in Pubmed: [24045663](https://pubmed.ncbi.nlm.nih.gov/24045663/).
- Chau I, Joulain F, Iqbal SU, et al. A VELOUR post hoc subset analysis: prognostic groups and treatment outcomes in patients with metastatic colorectal cancer treated with aflibercept and FOLFIRI. *BMC Cancer.* 2014; 14: 605, doi: [10.1186/1471-2407-14-605](https://doi.org/10.1186/1471-2407-14-605), indexed in Pubmed: [25142418](https://pubmed.ncbi.nlm.nih.gov/25142418/).
- Saltz LB, Clarke S, Diaz-Rubio E, et al. Bevacizumab in combination with oxaliplatin-based chemotherapy as first-line therapy in metastatic colorectal cancer: a randomized phase III study. *J Clin Oncol.* 2008; 26(12): 2013–2019, doi: [10.1200/JCO.2007.14.9930](https://doi.org/10.1200/JCO.2007.14.9930), indexed in Pubmed: [18421054](https://pubmed.ncbi.nlm.nih.gov/18421054/).
- Cherny NI, Sullivan R, Dafni U, et al. A standardised, generic, validated approach to stratify the magnitude of clinical benefit that can be anticipated from anti-cancer therapies: the European Society for Medical Oncology Magnitude of Clinical Benefit Scale (ESMO-MCBS). *Ann Oncol.* 2015; 26(8): 1547–1573, doi: [10.1093/annonc/mdv249](https://doi.org/10.1093/annonc/mdv249), indexed in Pubmed: [26026162](https://pubmed.ncbi.nlm.nih.gov/26026162/).
- Del Paggio JC, Azariah B, Sullivan R, et al. Do Contemporary Randomized Controlled Trials Meet ESMO Thresholds for Meaningful Clinical Benefit? *Ann Oncol.* 2017; 28(1): 157–162, doi: [10.1093/annonc/mdw538](https://doi.org/10.1093/annonc/mdw538), indexed in Pubmed: [27742650](https://pubmed.ncbi.nlm.nih.gov/27742650/).