



dr Tomasz Bandurski

Uzyskał I stopień specjalizacji z radiologii oraz tytuł specjalisty medycyny nuklearnej. Tematem jego rozprawy doktorskiej była *Ocena przydatności diagnostycznej scyntymammografii – porównanie z klasyczną mammografią rentgenowską i sonomammografią*. Pracuje w Zakładzie Informatyki Radiologicznej i Statystyki, gdzie do roku 2013 pełni obowiązki kierownika. Od chwili utworzenia w ramach Gdańskiego Uniwersytetu Medycznego Wydziału Nauk o Zdrowiu prowadzi zajęcia ze statystyki dla większości kierunków studiów na tym wydziale.

Biostatystyka i jej ograniczenia

lek. Łukasz Budyńko: Chcielibyśmy porozmawiać o statystyce medycznej. Interesują nas praktyczne wskazówki: narzędzia i materiały godne polecenia osobom, które zamierzają pogłębić wiedzę z tej dziedziny.

dr Tomasz Bandurski: Studia lekarskie nie kończą się egzaminem dyplomowym. Zdaje się wymagane egzaminy i uzyskuje się automatycznie tytuł, nie napisawszy żadnej pracy. Tymczasem pierwszy prawdziwy kontakt ze statystyką daje właśnie próba samodzielnego stworzenia artykułu naukowego. Wtedy nie tylko trzeba wykazać się umiejętnością czytania ze zrozumieniem artykułów naukowych – ja to nazywam bierną znajomością przedmiotu – ale również wiedzieć, jak czegoś dowieść samemu.

ŁB: Pan również ukończył kierunek lekarski. Jak Panu się udało uzupełnić swoją wiedzę?

TB: Muszę powiedzieć, że mój pierwszy kontakt ze statystyką miał miejsce na 4. roku wydziału lekarskiego. Był to tygodniowy blok zajęć i po wiem szczerze – nie uznawaliśmy go wówczas za niezbędny. Był to jeszcze jeden z pobocznych kursów, abstrakcyjnych, poza kanonem przedmiotów zawodowych. Potraktowaliśmy go wtedy jako obowiązek, który trzeba odbyć i tyle. Kiedy po studiach pojawiła się konieczność napisania własnej pracy, musiałem powrócić do tamtych źródeł.

ŁB: Większość lekarzy traktuje kontakt ze statystyką jako zło konieczne. Dlaczego Pan zajął się tym zagadnieniem?

TB: Zdałem sobie sprawę, że chcę naprawdę rozumieć to co robię. Właściwie wszystko zaczęło się od przygotowania doktoratu w Zakładzie Medycyny Nuklearnej. Doktorat był związany z oceną przydatności nowej metody diagnostycznej – scyntytmammografii. Do wykonania analizy potrzebowałem nietypowych narzędzi statystycznych. Stosowana przeze mnie analiza ROC, którą wykonywano zwykle ręcznie, była bardzo żmudna. Użycie komputera, który wykonywał ją w ciągu sekund, stało się przełomem.

ŁB: Obecnie z jakiej aplikacji Pan korzysta ?

TB: Używam głównie pakietu STATISTICA. Uczelnia ma licencję na ten program, wobec tego jest on łatwo dostępny i jednocześnie, w mojej opinii, dość prosty.

ŁB: Jakie książki Pan poleciłby człowiekowi, który chce ogarnąć w podstawowym stopniu możliwości nowoczesnych pakietów statystycznych?

TB: Nawiążę do pytania, dlaczego akurat wybrałem pakiet STATISTICA. Otóż w polskim piśmiennictwie istnieje bardzo dobry podręcznik użytkownika autorstwa Andrzeja Stanisza. Jest on dedykowany lekarzom i pracownikom ochrony zdrowia. *Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny* to dopracowany i łatwy do zrozumienia przewodnik dla osób, które nie mają podstaw matematycznych statystyki.

Zawodowi statystycy zaczynają studia od podstaw: matematyki, analizy matematycznej, probablistyki. Tam wiedza statystyczna jest niejako ćwiczeniem praktycznym z tych przedmiotów. Lekarze przyjmują zaś elementy statystyki na wiarę, opis metodologiczny traktując ją jako zło konieczne. Stosują statystykę w sposób dogmatyczny – trzeba wykonać konkretne czynności, bo tak musi być, ponieważ tak zrobiono w innym artykule, itp. W większości nie jest to postępowanie szkodliwe, chociażby dlatego, że istnieje pewna liczba rutynowych schematów działania. Jeżeli jednak chce się być krok przed konkurencją, trzeba wrócić do podstaw. Wtedy taka dogmatyczna wiedza stanowi ogromne obciążenie.

Przemysław Waszak: Książka Stanisza jest książką przeznaczoną dla osób w podobnej sytuacji...

TB: Zgadza się. Rzeczywiście, wypracował on metodę nauczania, która pozwala uniknąć początkowych pułapek wynikających z nieznamomości matematyki. Umożliwia w miarę swobodne poruszanie się po świecie statystyki. Osoby chcące dalej pogłębić swoją wiedzę z tego zakresu są stopniowo oswajane z językiem matematycznym.

ŁB: W drugim i trzecim tomie Stanisław wprowadza macierze, operatory...

TB: Tak, w kolejnych tomach ta tematyka jest już rozwinięta. Myślę, że podręcznik stanowi miękki start, godny polecenia dla lekarzy. Na przykład inne pozycje, które również odwołują się do medycyny, niestety narzucają bardzo stromą krzywą nauczania, uruchamiając od razu cały aparat matematyczny.

ŁB: Na ile młody badacz powinien rozwijać swoją wiedzę z matematyki? Na rynku istnieją książki w rodzaju *Matematyka w naukach przyrodniczych*...

TB: Ciężko powiedzieć, będzie to wynikało ze specyfiki tematu, w którą się zaangażuje. Sam niejednokrotnie wracałem do podstaw, chociażby do macierzy. Musiałem przypomnieć sobie szkolną wiedzę z tego zakresu, żeby następnie zrozumieć zasady budowania metod statystycznych. Natomiast nie ma przymusu zapisywania się z tego powodu na kolejne studia. Nie twierdzą, że każdy lekarz ma zostać matematykiem. Natomiast ważne jest uświadomienie sobie szerszego kontekstu analizowanych zagadnień. Lekarz powinien posiadać umiejętność rozmowy ze statystykiem, który z kolei musi być otwarty na zrozumienie problemów medycznych.

PW: Jak wiedza statystyczna przydaje się w pracy klinicznej lekarza?

TB: W pracy *stricte* klinicznej nie przydaje się zbyt często. Jeśli jednak analizujemy doniesienie naukowe, wówczas znając statystykę na pewno dostrzeżemy więcej niż tylko to, co jego autorzy uznali za stosowne zamieścić we wnioskach. Niektórzy twierdzą, że statystyka pozwala dopasować dane do zamierzonych rezultatów. Nie jest to może do końca prawda, ale patrząc na pracę, musimy być w stanie samodzielnie ocenić jej jakość. W takim rozumieniu statystyka jest przydatna każdemu lekarzowi.

PW: Mając przed sobą artykuł naukowy, co robi Pan, aby ocenić rzetelność artykułu?

TB: Szczerze mówiąc zwykle z braku czasu robię uproszczoną analizę. Czytam tytuł, porównuję z nim wnioski, następnie patrzę na opis metody i podejmuję decyzję, czy warto wejść w to głębiej. Ponadto stosuję również skalę Jadada tj. szereg pytań o sposób randomizacji, zaślepienie próby czy utratę chorych z badania.

ŁB: Jakie są najczęstsze błędy lekarze popełniane podczas analizy statystycznej?

TB: Błędy widać na bardzo różnych poziomach. Zasadnicza przyczyna tkwi w rozpowszechnieniu programów statystycznych. Interfejs graficzny użytkownika jest bardzo przystępny. Tworzy to mylne wrażenie, że skoro wiemy, gdzie kliknąć, to znaczy, że potrafimy bezbłędnie wykonać analizę statystyczną. Uzyskujemy w ten sposób jakieś wyniki, w których część wartości p jest poniżej 0,05. Oznaczałoby to istotność statystyczną. Często mechanizm ten doprowadza do wyciągania wniosków, które wydają się ciekawe, a są zupełnie bezzasadne.

Kiedyś, umiejętność liczenia testów statystycznych była mechanizmem samoweryfikującym. Umiały je wykonywać tylko osoby z odpowiednimi kompetencjami. W tej chwili każdy klika w przypadkowe funkcje w programie i coś generuje. Proszę mi wierzyć, widziałem nie raz artykuły z wnioskami zupełnie wyssanymi z palca. Ludzie powołują się na mądre testy, prezentują złożone wykresy o zupełnie absurdalnej treści. Chciałbym kiedyś porozmawiać z takim autorem...

ŁB: Zdarza się to w impactowanych czasopismach?

TB: Nie, poziom statystyki w impactowanych czasopismach oceniam dość dobrze. Z drugiej strony widać jej pewną dogmatyzację. Typowym przykładem jest przekonanie, żeby zawsze, przy rozkładzie normalnym dla cechy ilościowej używać średniej, a jeśli nie ma rozkładu normalnego – mediany. Jest to reguła bardzo ściśle przestrzegana. Ja się z tym nie do końca zgadzam. Uważam, że mediana jest dość mocno przecenianą miarą tendencji centralnej. Wykorzystuje się ją przy braku innej możliwości, ale czasem istnieją lepsze rozwiązania. Tymczasem użycie średniej przy braku normalności rozkładu jest traktowane jako grzech śmiertelny.

PW: Jakie rzeczy trzeba uwzględnić na etapie projektu badania, aby potem nie mieć trudności z wykonaniem statystyki?

TB: Statystyka pozwala nam zrozumieć dane tylko przy założeniu, że dobrze one reprezentują populację generalną. Należy pamiętać, że dane zbieramy z zamiarem poddania ich analizie statystycznej. Statystyka nie jest jakimś bytem odrębnym, tylko nieodłączną częścią dochodzenia naukowego. Myślmy o niej od początku, starając się w prawidłowy sposób dobrać badaną grupę. Zebrane dane muszą być reprezentatywne. To jest główny mankament większości prac, niezależnie od tego, czy z czasopism impactowanych, czy też nie. Kardynalnym błędem jest ignorowanie faktu, że pracujemy zawsze na wyselekcjonowanej grupie pacjentów. Do klinik nie trafiają ludzie bezpośrednio "z ulicy". To chorzy, którzy przeszli przez pewną selekcję. Nie da się bezpośrednio przełożyć wyników pracy badawczej na życie przeciętnego człowieka.

PW: Przykładowo: interesuje nas nadciśnienie tętnicze. Jeżeli chcemy wyciągnąć wnioski dotyczące populacji, popełnimy błąd, gdy analizujemy tylko ludzi z nadciśnieniem trafiających na oddziały szpitalne. Przebywają tam najcięższe przypadki.

TB: Zgadza się.

ŁB: Czy badania obserwacyjne budzą Pana nieufność? Ich konstrukcja jest z natury bardziej zawodna w porównaniu do eksperymentu klinicznego.

TB: Myślę, że nie można odrzucać badań obserwacyjnych. Wydaje mi się, że jeżeli w metodzie naukowej opisano: metodologię zdobywania danych, kontrolę jakości oraz kryteria brzegowe, wówczas jest to metoda, której można ufać.

Dużym problemem jest utrudniony dostęp do danych surowych. Trudno jest badaczowi udowodnić, że źle wykonał analizę, jeżeli każdy chroni swoje dane. Popularność zyskują natomiast szpitalne systemy informatyczne, które umożliwiają szerszemu gronu dostęp do pewnego rodzaju informacji, jak chociażby wyników dodatkowych. Jeszcze parę lat temu o dane trzeba było prosić poszczególne kliniki, co spotykało się na ogół z niechętną odpowiedzią. Rzecz podstawowa w naukach ścisłych, czyli możliwość weryfikacji wyników, w medycynie nie do końca działa.

ŁB: Młodzi badacze produkują wielką ilość badań korelacyjnych, przy czym beztrzesko przeskakują od stwierdzenia korelacji do przyczynowości.

TB: Programy statystyczne zachęcają do takiego postępowania z danymi. Zbieramy dowolne dane, wrzucamy je do macierzy i uzyskujemy korelację wszystkiego ze wszystkim. To, co podświetli się na czerwono, używamy do dalszej analizy. Wiąże się to ze wspomnianym wcześniej bezkrytycznym korzystaniem z programów statystycznych. Komputer sprowadzamy do roli czarnej skrzynki, do której wrzuca się maksymalną ilość danych i odbiera gotowy wynik.

Korelacje to ulubiony skrót myślowy młodych naukowców. Osoby niedoświadczone mają wrażenie, że są one równoważne dowodzeniu przyczynowości. Kiedyś był zwyczaj publicznego prezentowania w okresie bożonarodzeniowym śmiesznych doniesień, często opierających się na dziwnych korelacjach. I tak liczba gniazd bocianich korelowała z liczbą nowonarodzonych dzieci, z czego oczywiście krok od wniosku, skąd się one biorą na świecie.

PW: Przejdźmy do różnic między istotnością statystyczną i kliniczną. Na czym one polegają?

TB: Istotność statystyczna to zdjęcie błędu losowego z obserwowanych próbek. Analizując dane, nie wnikamy w sposób, w jaki one zostały zdobyte. Metodami matematycznymi szacujemy za to błąd losowy – na podstawie ich zmienności, odchylenia standardowego czy wariancji. W ten sposób oceniamy ich reprezentatywność, tzn. na ile uzyskany wynik jest dziełem przypadku. Przy małym błędzie mówimy o istotności statystycznej. Istotność kliniczna jest pojęciem szerszym, związanym nie tylko z doborem próby, ale i celem badania.

ŁB: Parametr, który od czasu wprowadzenia pakietów statystycznych zrobił furorę, to poziom istotności p. Wiele osób jest przekonanych, że przekroczenie sakramentalnej wartości $p < 0,05$ da każdemu twierdzeniu gwarancję prawdziwości. Jak to jest w istocie?

TB: Kiedyś każdą hipotezę rozpisywano na hipotezę zerową i hipotezę alternatywną. Aktualnie czynności te wykonują pakiety statystyczne, ujawniając zaledwie wynikową wartość p. Gdzieś umknął kontekst, gdzie i dlaczego to „p” się pojawia. Hipoteza zerowa jest pewnym rozwiązaniem naszego problemu badawczego. Możemy je przyjąć albo odrzucić.

Naszą decyzję w tej sprawie często porównuje się do orzeczenia sądu. Sąd może wydać błędną decyzję w dwóch wypadkach. Poważniejszym błędem byłoby jednak uznanie niewinnego za winnego, niż uznanie winnego za osobę niewinną [co ma szansę wyjść na jaw przy okazji kolejnego procesu – przyp. ŁB]. Przyjęcie błędnej hipotezy nazywamy błędem pierwszego rodzaju, odrzucenie prawdziwej – błędem drugiego rodzaju. Większym błędem jest przyjęcie fałszywej hipotezy. Tego staramy się unikać przede wszystkim...

ŁB: „Egzekucji” niewinnego?

TB: Tak. Prawdopodobieństwo popełnienia błędu pierwszego rodzaju jest właśnie naszym „ p ”. Od razu mamy odpowiedź na pytanie, co się dzieje z błędem drugiego rodzaju. Jeżeli skupimy się tylko na pierwszym, pomijamy istnienie drugiego. Ratuje nas to, że staramy się używać najmocniejszych testów statystycznych, na który pozwalają nam kanony.

Musimy jednak zdawać sobie zawsze sprawę, że błąd drugiego rodzaju jest tak samo realny. Niekiedy brak znamionnego efektu, przy wysokiej wartości p , skłania do wyciągnięcia wniosku, że analiza się nie powiodła. Najczęściej skutkuje to porzuceniem problemu. Tymczasem taki rezultat powinien dać nam również do myślenia. Może zwiększenie grupy badawczej, zmiana sposobu doboru grup lub wykonania kolejnych analiz zmieni ten stan rzeczy?

ŁB: W naukach przyrodniczych często przyjmuje się próg $p < 0,05$. Czy ma on uzasadnienie? W fizyce akceptowalna bywa dopiero istotność na poziomie 5-sigma.

TB: Tutaj mamy 2-sigma i mówimy już o sposobie znormalizowania rozkładu odchyłeń. Jeżeli miałbym postawić jakąś poprzeczkę, sądzę, że 0,05 wystarcza w naukach przyrodniczych i społecznych. Jest to zgodne z intuicyjnym postrzeganiem zjawisk w tych dziedzinach. Wszystko, co o nich wiemy, jest skutkiem obserwacji.

Aby zrozumieć, czym jest wartość p oraz poziom istotności, przytaczam na zajęciach następujący przykład. Należy rozstrzygnąć czy kostka, na której wypada cały czas 6 jest oszukana. Studenci zaczynają poszukiwać obiektywnych wykładników prawdopodobieństwa wypadania samych 6 w kostce nieoszukanej. Osoba może losować same 6, ponieważ ma zwyczajnie pecha. Pytam więc grupę, po ilu rzutach stwierdzi, że kostka jest jednak oszukana. Istnieje konieczność pewnej obiektywizacji wyników.

Trzeba wyznaczyć prawdopodobieństwo wyrzucenia określonej liczby szóstek pod rząd. W pojedynczym rzucie to $1/6$, w dwóch $1/36$, itp. Mówię wtedy, że to jest właśnie nasze „p”: malejące prawdopodobieństwo, że kostka jest nieoszukana, gdy wyniki eksperymentu zdają się temu przeczyć.

Przy nieustannym wypadaniu szóstek sprawa jest prosta. W rzeczywistości osiągnięcie modelowego poziomu p jest znacznie trudniejsze. To $0,05$ mówi nam, że coś jest na rzeczy, natomiast nie należy traktować tego jako absolutną prawdę. Jeżeli będziemy budować bardziej skomplikowane modele, np. wieloczynnikowe, możemy sobie pozwolić na podwyższenie albo obniżenie wartości progowej.

ŁB: Wspominał Pan, że analizy z wykorzystaniem skomplikowanych macierzy wprowadzają więcej zamętu niż pożytku. Czy warto zatem w ogóle korzystać z zaawansowanych technik przetwarzania danych?

TB: Oczywiście. Chciałem podkreślić, aby nie rozpoczynać zmagania ze statystyką od zagadnień najtrudniejszych i najbardziej wyszukanych. Nie traktujemy programu statystycznego jako przysłowiowej czarnej skrzynki, którą można wypełnić w dowolny sposób. Badacze mają obecnie dostęp do olbrzymich zbiorów danych. Ich ręczne przetwarzanie jest niemożliwe. W takich przypadkach zaawansowane techniki statystyczne mają rację bytu.

Próba integracji, wykorzystania istniejących wyników są metaanalizy. Uciekamy się do nich, kiedy nie ma możliwości stworzenia jednorodnej bazy danych. Najprościej byłoby oczywiście zebrać wszystkie dane składowe od poszczególnych badaczy, sprawdzając po drodze ich rzetelność, połączyć w jedno i zwyczajnie podsumować. Tego zrobić niestety się nie da, gdyż ludzie chronią swoje dane, zabraniają do nich dostępu. Stosujemy więc wyrafinowane narzędzia pomocnicze.

PW: Pana zdaniem jest to wiarygodne?

TB: Zależy od jakości badań składowych. Kalkulacja oparta wyłącznie na wynikach cudzych obliczeń nigdy nie będzie tak precyzyjna jak analiza uwzględniająca wszystkie dane. Kiedyś ludzie częściej zajmowali się naukami biomedycznymi z powołania. W momencie, kiedy za badaniami klinicznymi pojawiły się perspektywy korzyści finansowych, badaczom przestało zależeć na pilnowaniu poprawności metodologicznej. Stąd próby

zobiektywizowania wyników ich pracy w postaci EBM. Jest to odfiltrowanie zakłóceń od informacji. Każdy filtr może jednak sam wprowadzać pewne zniekształcenia.

ŁB: W naukach medycznych najczęściej operujemy prawdopodobieństwem obiektywnym, którego wyrazem jest „p”. Mało mówi się o prawdopodobieństwie subiektywnym, związanym z twierdzeniem Bayesa. Czy warto się tym zainteresować?

TB: Uważam, że jest to bardzo ciekawe. Ze zmianami względnymi prawdopodobieństwa zetknąłem się podczas pracy nad doktoratem. Ważny jest sposób konstrukcji modelu matematycznego. Im prostszy model przyjmemy, tym prostsze wzory do obliczenia uzyskamy. Jednak wraz ze zbliżaniem się do rzeczywistości, musimy uciekać się do metod coraz bardziej skomplikowanych.