

Dr hab. Robert K. Filipkowski

Środowiskowe Laboratorium Behawioralno-Metaboliczne

Instytut Medycyny Doświadczalnej i Klinicznej im. M. Mossakowskiego PAN w Warszawie

Medycyna translacyjna i choroba Alzheimer'a; wybrane potencjalne przyczyny niepowodzeń

„Our job as scientists is to discover truths about the world.”

(Joseph P. Simmons, Leif D. Nelson, Uri Simonsohn)

„Krąg IV: polujący na wartość p . Tu znaleźli się ci, którzy wypróbowali każdy test statystyczny, o którym tylko zdołali wyczytać, póki nie uzyskali wartości p mniejszej niż 0,05. Krąg ten stanowi przeogromne jezioro o mętnej wodzie. Grzesznicy siedzą w łodziach i aby jeść, muszą łowić ryby. Na szczęście dysponują olbrzymim zestawem wędek i podbieraków – produktami firm Bayes, Student, Spearman; i wielu innych. Niestety, jedynie jedna na 20 ryb nadaje się do spożycia, więc grzesznicy tego kręgu są nieustannie głodni.”

(Neuroskeptic¹. *The Nine Circles of Scientific Hell*)

W ramach badań przedklinicznych dotyczących choroby Alzheimer'a popularne jest wykorzystanie myszy jako modelu choroby i testów behawioralnych mierzących uczenie się i pamięć u badanych zwierząt np. basenu Morrisa². Trafność (*ang.* validity) tych modeli jest dyskusyjna i dyskutowana. Jeden z problemów to fakt, że stosujemy zwierzęta z mutacjami spotykanymi w rodzinnych wariantach choroby, podczas gdy podstawowy problem zdrowotny i społeczny stanowi jej forma spontaniczna. Spektakularnym niepowodzeniem zakończyły się próby kliniczne

¹ <https://en.wikipedia.org/wiki/Neuroskeptic>; neuroskeptic@gmail.com, <http://blogs.discovermagazine.com/neuroskeptic/>

² Labirynt wodny Morrisa / basen Morrisa to behawioralny model badania pamięci deklaratywnej i przestrzennej u gryzoni, który stanowi przykład warunkowania instrumentalnego (skinnerowskiego). Jest to duży (np. 150 cm średnicy dla myszy, 180 cm – dla szczurów), okrągły basen wypełniony wodą. Tuż pod jej powierzchnią, w określonym, stałym miejscu ukryta jest platforma. Gdy mysz lub szczur zostają umieszczone w wodzie na np. 60 s, w jednym z czterech różnych miejsc startowych, zaczynają aktywnie pływać, szukając wyjścia z awersyjnej sytuacji. Jeśli trafią na platformę, z reguły wdrapują się na nią. Z próby na próbę uczą się, że platforma jest miejscem ucieczki i gdzie się ona znajduje w przestrzeni. Kierują się przy tym kształtem pomieszczenia i znajdującymi się w nim obiektami, tworząc mapę całego pokoju. Z dnia na dzień, z próby na próbę zwierzęta odnajdują platformę szybciej a pamięć o jej położeniu utrzymuje się przez wiele dni. Test pamięci in. próba testowa (*ang.* probe trial) polega na wyjęciu platformy i ponownym umieszczeniu gryzonia w wodzie na np. 60 s. „Nauczona/pamiętająca” mysz spędza wtedy najwięcej czasu w tej wirtualnej ćwiartce basenu, gdzie znajdowała się platforma; wielokrotnie także przepływając przez zajmowane przez nią miejsce. Miarą pamięci jest więc czas spędzony w docelowej ćwiartce (*ang.* target quadrant) i liczba przecięć platformy (*ang.* crossings). Do właściwego wykonania tego zadania niezbędna jest obecność nienaruszonej formacji hipokampa. Wielokrotnie pokazywano zaburzenia pamięci w basenie Morrisa u zwierząt będących modelami choroby Alzheimer'a.

zastosowania szczepionki skierowanej przeciw beta-amyloidowi. Próby te wypadły wcześniej pomyślnie w badaniach przedklinicznych z udziałem myszy. Przyczyn zaistniałej sytuacji można szukać (między innymi!) właśnie w braku odpowiedniego zwierzęcego modelu choroby, który może wynikać z naszej niedostatecznej wiedzy o jej etiologii, ze złożoności ludzkiego mózgu – w porównaniu z mózgiem mysim, z odmiennego, lokalnego, selektywnego postępu choroby u ludzi, z krótkiego życia organizmów modelowych i... błędów metodycznych. Na pewnym aspekcie tych ostatnich chciałbym się poniżej skupić.

Jako zawodowi naukowcy poddani jesteśmy nieustannej presji zewnętrznej (szefów, pracodawców i grantodawców) i wewnętrznej (konkurencja wewnątrzgatunkowa występująca pomiędzy naukowcami), by coraz lepiej i coraz więcej publikować. Jednocześnie mamy ogromną swobodę w podejmowaniu wielu decyzji dotyczących np. sposobu prowadzenia doświadczeń. Wraz z ogólnym kryzysem wartości i pojęć prowadzi to do sytuacji, w której naczelną wartością dla naukowca nie jest poszukiwanie prawdy, ale uzyskanie małego p mniejszego od 0,05. Celem staje się uzyskanie wyniku znamiennego statystycznie, a raczej – uzyskanie znamienności statystycznej otrzymanego wyniku. To wszystko z kolei zgodnie z *efektem Lucyfera* (Zimbardo 2008) powoduje u badaczy stopniowe schodzenie z drogi właściwej praktyki naukowej i powolne pograżanie się w kolejnych kręgach naukowego piekła (Neuroskeptic 2012) i prowadzi do niezauważania drobnych dowolności metodycznych, zachęcania do niewielkich nadużyć proceduralnych, wyolbrzymiania znaczenia własnych odkryć, snucia naciąganej naukowej narracji o rzekomym precyzyjnym planowaniu owych odkryć, polowania na wartość p (patrz motto II), odrzucania „niewygodnych” wyników, publikowania wybiórczego a nawet plagiatów i oszustw.

Problem polega m.in. na tym, że nasze zwykłe, codzienne, rutynowe działania prowadzą do tego, że łatwość z jaką można opublikować „znamienne statystycznie” wyniki potwierdzające jakąkolwiek hipotezę, osiągnęły poziom nieakceptowalny (Simmons, Nelson, Simonsohn 2011). Problem przebił się do świadomości psychologów społecznych (zachęcam do lektury artykułu Susan Dominus, 2017!), jednak dopiero powoli dociera do świadomości biologów doświadczalnych.

Problem opisali Joseph P. Simmons, Leif D. Nelson i Uri Simonsohn (2011) i określili jako zbyt liczne *stopnie swobody naukowca* (ang. researcher degrees of freedom). Pokusa poprawiania i ulepszania (czyt. fałszowania) wyników nie jest obca wszystkim badaczom, w tym autorowi, jednak w tym podsumowaniu chciałbym skupić się na badaczach zachowania zwierząt i pokazać, jak wyniki i postulaty ww. trójki badaczy odnoszą się do naszego pola badań. Autorzy ci zauważają (a autor niniejszego podsumowania dodaje szczegóły), że „swobodny naukowiec” stosuje szeroką gamę „chwyków” (czyt. drobnych oszustw), co prowadzi do sytuacji, w której możemy potwierdzić dowolną tezę.

„Swobodny” naukowiec zbiera dane:

- przeprowadzając wiele różnych porównań tych samych grup doświadczalnych, np. kontrolnej i eksperymentalnej; prawdopodobieństwo, że co najmniej jedna z wielu takich analiz, prowadzonych na poziomie 5%, obarczona jest błędem pierwszego stopnia (*ang.* false positive) jest z konieczności większe niż 5%,
- zbierając dane doświadczalne do momentu otrzymania zadawalającego wyniku, tzn. dobierając dane i sprawdzając, czy wyniki uzyskane w badaniu w sposób odpowiedni i znamieny „się różnią”; im więcej etapów, na których znamienność różnic jest sprawdzana, tym więcej stopni swobody,
- zbierając dane doświadczalne na podstawie badania wielu osobników, w tym także „warunkowo” takich, które nie w pełni spełniają założenia metodyczne doświadczenia, np. są zbyt stare lub pobudzone, by w ostatecznej analizie uwzględnić te dane lub z nich zrezygnować w zależności od tego, „co wyjdzie”; im więcej takich grup i osobników, tym więcej stopni swobody,
- wykluczając wyniki odstające (*ang.* outliers) na różne preferowane/ulubione sposoby; im więcej takich „sposobów”, tym więcej stopni swobody.

„Swobodny” naukowiec analizuje dane:

- analizując uzyskane wyniki na wiele sposobów, wiele uzyskanych wielkości ze sobą porównując, np. analizując wyniki z basenu Morrisa: czas (w %) w ćwiartce docelowej uzyskany przez osobniki z grupy eksperymentalnej w porównaniu z 25% (poziom przypadkowości), w porównaniu z wynikiem uzyskanym przez grupę kontrolną w ćwiartce docelowej, w innych ćwiartkach (wszystkich), średnio we wszystkich ćwiartkach, w ćwiartkach przylegających do ćwiartki docelowej, średnio w ćwiartkach przylegających do ćwiartki docelowej, biorąc pod uwagę całą próbę, lub tylko jej część np. od momentu, gdy zwierzę zaczęło aktywnie/ochoczo pływać lub gdy wypłynęło z ćwiartki naprzeciwległej (*ang.* opposite) lub generalnie, gdy aktywnie pływało itp.,
- analizując liczbę przecięć miejsca zajmowanego przez platformę lub miejsca trochę większego lub dwa razy większego itp.,
- łącząc wybrane grupy doświadczalne, porównując wybrane grupy doświadczalne,
- wybierając jedynie te zmienne zależne spośród wielu badanych, które „wyszły”,
- analizując w sposób kreatywny i *ad hoc* wyodrębnione „śródogrupy”; np. wybierając z puli wyników jedynie wyniki samców, osobników w określonym wieku itp.,
- stosując transformację lub nie stosując transformacji danych w celu ich analizy testami parametrycznymi lub nieparametrycznymi, analizami wariancji kilku stopni, po sprawdzeniu lub niesprawdzeniu spełniania założeń testu statystycznego przez uzyskane wyniki, by zakończyć analizę testem typu *post-hoc* wybranym spośród szerokiej gamy takich testów,

- by, w razie niepowodzenia (np. $p = 0,057$), zastosować jedno z eufemistycznych określeń zastępujących brak znamienności statystycznej na poziomie 0,05 (np. „weakly significant”, „almost significant” i inne z ponad 500 odnotowanych przez autora wpisu – Matthew Hankinsa, 2013 – wariantów)
- lub/i opublikować jedynie tę część wyników, która „wyszła”.

Przeprowadzone przez autorów symulacje pokazały, że gdy zastosowanych kombinacji (stopni swobody) będzie odpowiednio wiele (np. dwie zmienne zależne; dodatkowe 10 pomiarów – jeśli początkowe 20 pomiarów dało wynik nieznamienny; możliwość wyróżnienia ko-wariantu np. płci; założenie trzech różnych warunków z możliwością rezygnacji z jednego z nich), szansa na uzyskanie wyniku typu *false-positive* przy $p < 0,05$ wzrasta do ponad 60%. W ten sposób Simmons, Nelson i Simonsohn (2011) pokazali na przykład, że osoby po wysłuchaniu piosenki opowiadającej o sytuacji osób w starszym wieku („When I’m Sixty-Four” Beatlesów) stawały się o niemal półtora roku młodsze w porównaniu z osobami z grupy kontrolnej, które słuchały piosenki „Kalimba”.

Nie wiadomo, na ile stosowanie wyżej wymienionych „sposobów” przyczyniło się do opublikowania fałszywych wyników dotyczących działania szczepionki przeciw beta-amyloidowi u myszy transgenicznych. Niemniej na przykład w pracy Janusa i wsp. (2000) pokazano, że szczepienie amyloidem prowadzi do poprawy funkcji kognitywnych u myszy, podczas gdy omawiane różnice pomiędzy badanymi grupami były niewielkie, znamienność statystyczna pokazanych różnic słaba a poziom uczenia się u myszy kontrolnych zmienny (Ryc. 2b), użyte grupy były nieliczne, np. $n = 5$, $n = 6$, n około 6 [sic], nie było grupy transgenicznej, która nie dostała szczepionki, różnice między grupą kontrolną i transgeniczną wystąpiły w 11. i 23. tygodniu życia (dla 15 i 19 tygodni nie wystąpiły!), nie pokazano wyników z próby testowej, w ramach której w ogóle nie stwierdzono różnic.

Co robić!? Jak żyć!?

Celem nas, naukowców, nie jest publikowanie jak największej liczby artykułów, ale odkrywanie i ogłaszanie prawdy (patrz motto I). Jednakże „osoba znajdująca się w szponach Systemu po prostu postępuje zgodnie z nim, czyniąc to, co wydaje jej się naturalnym sposobem reagowania w określonym miejscu i czasie (Zimbardo 2008; str. 202).” Jedyna nadzieja w jednostkowym wyborze, Twoim wyborze, w ramach którego nie potraktujesz powyższych informacji jako instrukcji obsługi. Pisz i mów prawdę (lub przynajmniej nie kłam!) (Peterson 2018; Reguła 8, str. 203). Opisz w publikacji to, co rzeczywiście zrobiłeś. Krzycz jako np. doktorant, że król (np. promotor) jest nagi. Nie bierz udziału w uporczywym udowadnianiu zerowej hipotezy. Szkoda naukowego życia i twojego miejsca w powstającej historii nauki. Łatwo w nim i w niej wylądować jako zwolennik rozwiązań błędnych i chybionych.

Literatura

1. Dominus S. When the Revolution Came for Amy Cuddy. The New York Times Magazine. Oct. 22, 2017, page 29 (<https://www.nytimes.com/2017/10/18/magazine/when-the-revolution-came-for-amy-cuddy.html>)
2. Hankins M. Still Not Significant. 2013.
<https://mchankins.wordpress.com/2013/04/21/still-not-significant-2/>
3. Janus C. i wsp. A β Peptide Immunization Reduces Behavioural Impairment and Plaques in a Model of Alzheimer's Disease. 2000. Nature. 408: 979-982.
4. Neuroskeptic. The Nine Circles of Scientific Hell. 2012. Perspectives on Psychological Science. 7: 643-644.
5. Simmons J.P., Nelson L.D., Simonsohn U. False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. 2011. Psychological Science. 22: 1359-1366.
6. Zimbardo P. Efekt Lucyfera. PWN Warszawa. 2008.
7. Peterson J.B. 12 Rules for Life, An Antidote for Chaos. Penguin Books. 2018.