Akceptuję
W ramach naszej witryny stosujemy pliki cookies w celu świadczenia państwu usług na najwyższym poziomie, w tym w sposób dostosowany do indywidualnych potrzeb. Korzystanie z witryny bez zmiany ustawień dotyczących cookies oznacza, że będą one zamieszczone w Państwa urządzeniu końcowym. Możecie Państwo dokonać w każdym czasie zmiany ustawień dotyczących cookies. Więcej szczegółów w naszej Polityce Prywatności



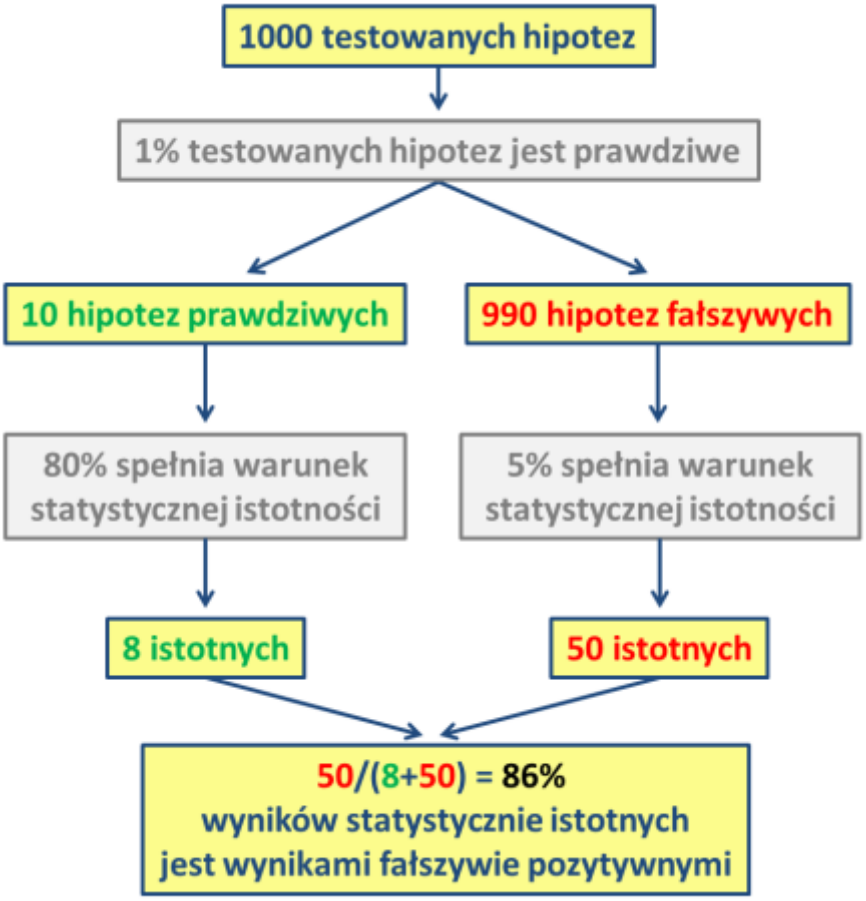
# Dlaczego nie wiadomo, czy większość publikowanych badań jest jednak prawdziwa?

**Dawno, dawno temu, w pierwszych latach istnienia Public Library of Science, profesor medycyny na Uniwersytecie Stanforda John Ioannidis opublikował w piśmie PLoS Medicine pracę, która stała się najbardziej popularną i cytowaną pracą zamieszczoną w tym periodyku, notując na swoich koncie w okolicy tysiąca cytowań, ponad 10 tysięcy udostępnień w mediach społecznych i Altmetric Score 896 (i ciągle rośnie).**



O czym była ta praca? Otóż był to esej o dźwięcznej nazwie „Why Most Published Research Findings Are False“, w którym, jak się łatwo domyślić, Ioannidis dowodził, że w przypadku znakomitej większości publikacji naukowych ich wyniki są fałszywe.

Ioannidis wyjaśnia, że w większości prac naukowych w dzisiejszych czasach o prawdziwości wyników orzeka się tylko na podstawie jednej metryki: tego, czy w pojedynczym studium udało się uzyskać wynik statystycznie istotny z formalnego punktu widzenia – tzn. czy znienawidzona wartość p była mniejsza niż, zazwyczaj, 0.05 (tu odsyłam do postu o naukowym piekle, a zwłaszcza do nonalogu badacza na jego końcu i przykazania czwartego). Badacz tłumaczy dalej, że prawdopodobieństwo prawdziwości wyników badań naukowych zależy także od tego, jaka jest szansa, że będą prawdziwe w pierwszej kolejności1, od mocy testu, i dopiero na końcu od poziomu statystycznej istotności, czyli wartości p. W skrócie, rozważania Ioannidesa sprowadzają się do tego, że testowanych hipotez, które są fałszywe, jest znacznie więcej niż prawdziwych i nawet jeśli badania weryfikują znacznie większy procent tych drugich, nawet niewielki ułamek wyników fałszywie pozytywnych jest wciąż znacznie większy niż liczba wyników prawdziwych.




Ten schemat (przerobiony na polszczyznę z cytowanej na końcu pracy Leeka) przedstawia schemat myślenia Ioannidesa.

Autor spisuje wnioski wynikające z opisywanego przez niego modelu: po pierwsze, im mniejsze badanie (mniej liczna próba), tym mniejsza szansa, że otrzymane wyniki są prawdziwe. Po drugie, im mniejszy jest obserwowany efekt, tym mniejsza szansa, że wyniki są prawdziwe. Po trzecie, im większa jest liczba i im mniej dokładna selekcja badanych relacji, tym mniejsza szansa, że wyniki są prawdziwe (zaczynacie już pewnie dostrzegać prawidłowość). Po czwarte, im większa elastyczność eksperymentalnych projektów, im większa elastyczność definicji, wyników, sposobów analizy, tym mniejsza szansa, że wyniki są prawdziwe. Po piąte, im większe zaangażowanie finansowe, światopoglądowe lub jeszcze inne w badania, tym mniejsza szansa, że wyniki są prawdziwe. I wreszcie po szóste, im bardziej seksowne jest pole badań i im więcej zespołów badawczych zajmuje się danym tematem, tym mniejsza szansa, że wyniki są prawdziwe.

Praca Ioannidisa wywołała, jak się można domyślić, sporo zamieszania. Nie to, żeby ktoś się przejął i zaczął zmieniać sposób prowadzenia badań biomedycznych. Ale jego esej wraca jak bumerang w dyskusjach za każdym razem, gdy ktoś chce dać przykład tego, jak niewydajna jest nauka (nawet wówczas, gdy badacze robią wszystko w porządku). W 2007 pojawiło się zresztą kilka publikacji, próbujących z tej pracy wyciągnąć jakieś bardziej praktyczne wnioski. Benjamin Djulbegovic i Iztok Hozo zaproponowali, także w PLoS Medicine, model, który ma nam pomóc zdecydować, które z tej większości najprawdopodobniej nieprawdziwych wyników są jednak użyteczne.

Bo nie ukrywajmy, nigdy nie jest i nie będzie tak, że nauka da nam ostateczną odpowiedź na jakieś pytanie. Zawsze można coś zmierzyć lepiej, dokładniej. Zawsze można coś poprawić, coś troszkę

inaczej zinterpretować. Skoro absolutnej prawdy nie da się uzyskać, powstaje pytanie: które z dostępnych wyników możemy zaakceptować, jako najlepsze z najgorszych. I na to pytanie Djulbegovic i Hozo próbowali odpowiedzieć.

Kolejna grupa badaczy opublikowała – w tymże samym piśmie, także w 2007 roku – pracę, w której dumali nad problemami przedstawionymi przez Ioannidesa. I doszli do wniosków niespecjalnie porażających, a mianowicie: powtórzenie eksperymentu zwiększa prawdopodobieństwo, że uzyskane wyniki są prawdziwe. Ameryki raczej tą publikacją nie odkryli, ale stali się z pewnością kolejnym głosem nawołującym do powtarzania już raz wykonanych i opisanych eksperymentów.

Bo przecież na tym polega podstawowy problem. Badania są publikowane dzisiaj na chybcika, często na przerażająco małych próbach. I z wielką szkodą dla nauki w ogóle: nierzadko opublikowanie wyniku, który de facto jest tylko tym tzw. proof-of-principle (czyli jakościową demonstracją, że jakieś zjawisko jest prawdopodobne), i za którym powinny iść zakrojone na szeroką skalę badania ilościowe, prowadzi do tego, że takich badań nie wykonuje się w ogóle. Bo skoro wynik został już opublikowany, to potwierdzenia żadne szanujące się pismo wysokoimpaktowe nie weźmie. A grantodawcy, wbrew regułom, i tak świadomie lub nie patrzeć będą milszym okiem na publikacje w Nature i Science, niż w Journal of Random and Not Very Spectacular Confirmatory Results.

No i gdy już całkiem wydawać by się mogło, że sytuacja jest tak pesymistycznie nierozwiązywalna, że bardziej się nie da; gdy już miałem pisać mail do Ioannidesa z krzykliwym tytułem „zabiłżeś pan inspirację do badań!", na serwery arxiv złożona została ku uciesze gawiedzi praca pod dźwięcznym tytułem „Empiryczne szacunki sugerują, że większość publikowanych badań medycznych jest prawdziwa".

Praca wyszła spod pióra Leah Jagera i Jeffa Leeka (którego niektórzy czytelnicy bloga będą być może pamiętać z tego wpisu). Panowie postanowili wziąć pod lupę jedno z podstawowych założeń pracy Ioannidesa: a mianowicie to, czy rzeczywiście odsetek wyników fałszywie pozytywnych jest aż tak wysoki. W tym celu dokonali karkołomnego przekopywania się przez abstrakty prac medycznych opublikowanych w kilku najlepszych medycznych periodykach naukowych (m.in. Lancet, BMJ, NEJM) w latach 2000 – 2010, w celu wyłowienia z nich wartości p.

Na podstawie tych danych skonstruowali model, który pozwolił im oszacować odsetek wyników fałszywie pozytywnych, a wartość, którą otrzymali, chociaż wciąż większa niż byśmy sobie życzyli, jest jednak znacznie, znacznie mniejsza od 86% Ioannidesa: Leek i Jager twierdzą mianowicie, że jest to raczej w okolicach 14%, co oznacza, że ich krzykliwy tytuł ma rację. Większość publikowanych wyników jednak jest prawdziwa.

W internecie już rozgorzała dyskusja nad tą pracą, a adwersarze przerzucają się statystycznymi argumentami wagi zbyt dużej na moją słabą głowę. Z tym, że trzeba tutaj zaznaczyć, że większość rozmówców chce wierzyć, że ten rezultat jest prawdziwy, bo chce wierzyć, że możemy ufać wynikom badań medycznych. Część z nich nie jest po prostu przekonana, czy model Leeka i Jagera rzeczywiście problem rozwiązuje.

Historia ma też kolejne dno: w ostatnich latach media naukowe obiegło kilka tekstów, które wsparły podstawę rozważań Ioannidesa. W 2011 Nature Reviews Drug Discovery opublikował pracę trójki badaczy z firmy Bayer, którzy donieśli o dość niepokojącym odkryciu. Otóż firmy farmaceutyczne wyniki prac opisujących odkrycia nowych leków weryfikują samodzielnie, zanim się w jakąś badawczą alejkę zaczną same pchać. Autorzy opisują wyniki takich prób weryfikacji dla 67 różnych projektów. Dane literaturowe zgadzały się z ich danymi w 21% przypadków. W dodatkowych 7% główne dane były takie same, ale pojawiły się rozbieżności w detalach. W 2/3 (słownie: dwóch

trzecich!) przypadków, wyniki były niepowtarzalne. Z kolei w marcu 2012 dwóch innych badaczy w komentarzu opublikowanym w Nature opisało badania naukowców w firmy Amgen, którzy usiłowali powtórzyć doświadczenia z 53 publikacji. Udało im się to w mizernych 11% przypadków.

Zatem nawet jeśli model Jagera i Leeka okaże się poprawny, może się okazać, że nijak nie pomoże to w odbudowaniu zaufania do badań biomedycznych. W końcu model jest tylko modelem. I znacznie bardziej do nas wszystkich niż model powinny przemawiać właśnie wyniki prób powtórzenia badań. Nie powinno zatem zapewne dziwić, że gdy Nature ogłaszał swoją listę 10 Naukowych Postaci 2012 Roku, na liście znalazła się Elizabeth Iorns, fundatorka Reproducibility Initiative, projektu, który skupia się dokładnie na tym: na weryfikacji wyników publikacji (prace zgłaszać mogą sami autorzy, zaś jeśli wyniki uda się zreplikować, to rezultaty publikowane są w PLoS ONE). Niestety nie mogę nigdzie znaleźć informacji na temat tego, jaki odsetek nadsyłanych prac udaje się rzeczywiście powtórzyć (ale też sam projekt jest bardzo młody, więc może za wcześnie na to).

Jeśli jest w tej całej historii jakiś pozytywny akcent, to może właśnie to, że pojawiają się takie inicjatywy jak ta dr Iorns, że dostrzega się i coraz częściej i głośniej mówi o takich problemach, jak niepowtarzalność wyników, że większość badaczy chce, aby ich wyniki poddano weryfikacji (70% autorów zaproszonych do nadesłania swoich prac do RI wyraziło w tej kwestii entuzjazm), co oznacza, że nawet jeśli nie uda się ich powtórzyć, to raczej nie z powodu złych chęci czy intencji samych badaczy. Na koniec warto może dodać, że RI wpisuje się bardzo pięknie w retorykę Ioannidesa. Bo nawet jeśli obserwowany efekt jest niewielki, nawet jeśli próbka jest niespecjalnie duża (czasem zresztą nie da się inaczej), jeśli ten sam rezultat obserwować będą coraz to i kolejne grupy, szansa, że wynik jednak jest prawdziwy, dramatycznie wzrośnie. A to końcu jest to, co chcemy wiedzieć.

Przypisy:

1. Przed badaniem. Bo łatwo sobie wyobrazić, że jeśli spróbujemy wykonywać eksperyment próbując dowieść, że nieprawdziwe są fundamentalne prawa fizyki – np. podrzucać jabłko do góry i pokazywać, że wcale nie spada ono potem na dół – i uzyskamy wartość $p<0.05$, to nie oznacza jeszcze, że udało nam się obalić teorię grawitacji!

2. Ioannidis, J. (2005). Why Most Published Research Findings Are False PLoS Medicine, 2 (8) DOI: 10.1371/journal.pmed.0020124

3. Djulbegovic, B., & Hozo, I. (2007). When Should Potentially False Research Findings Be Considered Acceptable? PLoS Medicine, 4 (2) DOI: 10.1371/journal.pmed.0040026

4. Moonesinghe, R., Khoury, M., & Janssens, A. (2007). Most Published Research Findings Are False—But a Little Replication Goes a Long Way PLoS Medicine, 4 (2) DOI: 10.1371/journal.pmed.0040028

5. Leah R. Jager, & Jeffrey T. Leek (2013). Empirical estimates suggest most published medical research is true Arxiv arXiv: 1301.3718v1

6. Prinz, F., Schlange, T., & Asadullah, K. (2011). Believe it or not: how much can we rely on published data on potential drug targets? Nature Reviews Drug Discovery, 10 (9), 712-712 DOI: 10.1038/nrd3439-c1

7. Begley, C., & Ellis, L. (2012). Drug development: Raise standards for preclinical cancer research Nature, 483 (7391), 531-533 DOI: 10.1038/483531a

*Źródło: http://nicprostszego.wordpress.com*

http://laboratoria.net/home/16495.html
**Informacje dnia:** Jazda na rolkach - Czy jest dobrym sportem? Polimer o właściwościach przeciwgrzybiczych Stypendia ministra nauki dla niemal 400 studentów Skuteczniejsze leczenie chorych na nowotwory krwi Tylko 36% transgranicznych wód podziemnych ma ochronę Technologia ultradźwiękowa w diagnostyce chorób Jazda na rolkach - Czy jest dobrym sportem? Polimer o właściwościach przeciwgrzybiczych Stypendia ministra nauki dla niemal 400 studentów Skuteczniejsze leczenie chorych na nowotwory krwi Tylko 36% transgranicznych wód podziemnych ma ochronę Technologia ultradźwiękowa w diagnostyce chorób Jazda na rolkach - Czy jest dobrym sportem? Polimer o właściwościach przeciwgrzybiczych Stypendia ministra nauki dla niemal 400 studentów Skuteczniejsze leczenie chorych na nowotwory krwi Tylko 36% transgranicznych wód podziemnych ma ochronę Technologia ultradźwiękowa w diagnostyce chorób

## Partnerzy