

MAREK KWIEK*, WOJCIECH ROSZKA**

Reguła 10/50.

Analiza dorobku 150 tysięcy polskich naukowców pokazuje trwałość systemowych nierówności w nauce

1. Wprowadzenie

W prezentowanej pracy analizujemy udział najbardziej produktywnych polskich naukowców (górny 1%, 5% i 10%) w krajowej produkcji publikacyjnej w ciągu ostatnich 30 lat (1992–2021). Badamy ich udział podłużnie, czyli w ujęciu czasu, i na mikropoziomie poszczególnych naukowców. Szczegóły ich indywidualnych historii publikacyjnych (pochodzące z jednostkowej bazy danych Scopus) są badane osobno, a następnie agregowane do wyższych poziomów: dyscyplin akademickich, typów instytucji, grup wieku akademickiego (doświadczenia publikacyjnego) i sześcioletnich okresów czasu. Najbardziej produktywni naukowcy są porównywani z resztą polskiej publikującej kadry naukowej, czyli z pozostałymi 99%, 95% i 90%.

Pokazujemy, że ich udział w krajowej produkcji publikacyjnej jest zaskakująco stabilny w czasie: chociaż nauka akademicka przechodziła bezprecedensowe zmiany, najbardziej produktywni naukowcy w latach 1992–2021 zawsze odgrywali strukturalnie podobną, fundamentalną rolę. Górny 1% naukowców odpowiadał średnio za 10% krajowej produkcji publikacyjnej (co można nazwać regułą 1/10), a górnych 10% naukowców odpowiadało za prawie 50% tej produkcji (co można nazwać regułą 10/50). I było tak przez 30 badanych lat.

Stosujemy podejście ilościowe i poddajemy analizie 152 043 polskich naukowców pochodzących z 15 dyscyplin STEM (nauki ścisłe, techniczne, inżynieryjne, matematyczne i medyczne) oraz ich 587 558 artykułów naukowych opublikowanych w pięciu sześcioletnich okresach (w latach 1992–2021) w czasopiśmie indeksowanych w bazie Scopus. Skupiamy się na naukowcach i ich cechach oraz na prezentowanych przez nich wzorcach publikacyjnych w ramach poszczególnych okresów. Nasze klasy najbardziej

* Prof. dr hab. Marek Kwiek (marek.kwiek@amu.edu.pl), Institute for Advanced Studies in Social Sciences and Humanities (IAS), Uniwersytet im. Adama Mickiewicza w Poznaniu; ORCID: orcid.org/0000-0001-7953-1063

** Dr Wojciech Roszka (wojciech.roszka@ue.poznan.pl), Uniwersytet Ekonomiczny w Poznaniu, Centrum Studiów nad Polityką Publiczną UAM w Poznaniu; ORCID: orcid.org/0000-0003-4383-3259

produktywnych naukowców są definiowane przy użyciu metadanych dotyczących publikacji i cytowań, a indywidualna produktywność publikacyjna jest normalizowana do sześcioletniego okresu, dyscypliny i prestiżu czasopism według ich miejsca w ujęciu rang percentylowych (miara CiteScore używana w bazie Scopus w przedziale 0–99).

Prezentowane badanie wykorzystuje podłużne (longitudinalne) ujęcie wzorców publikacyjnych, w którym analizowani są wszyscy aktywni badawczo polscy naukowcy obecni w bazie Scopus w dziedzinach STEMM. Na podstawie wyników naszych wcześniejszych badań ankietowych (Kwiek 2016; Kwiek 2018a) spodziewaliśmy się, że udział najbardziej produktywnych naukowców w całości krajowego dorobku publikacyjnego może być duży. W obu pracach reguła 10/50 pojawiła się po raz pierwszy.

Jednak nasze kompleksowe podejście podłużne nie było możliwe do zastosowania przed pojawieniem się dużych zbiorów danych (typu *structured Big Data*), czyli w naszym przypadku kompletnego nieprzetworzonego zbioru danych Scopus. Nasze wyniki nie odnoszą się do pojedynczego punktu w czasie (np. trzech lat, jak w większości badań ankietowych kadry akademickiej). Korzystając z danych panelowych, możemy badać długoterminową stabilność (lub zmiany) produktywności badawczej w Polsce. Upływ czasu ma duże znaczenie w badaniach karier naukowych, dlatego w naszej analizie wykorzystujemy kolejne, równe, sześcioletnie okresy.

W niemal wszystkich dotychczasowych analizach wysokiej produktywności badawczej brakowało właśnie komponentu czasowego (wyjątki, w których analizowano maksymalnie 10–12 lat obejmują trzy prace: Turner i Mairesse 2005; Kelchtermans i Veugelers 2013; oraz Abramo, D'Angelo i Soldatenkova 2017; szerzej Kwiek i Roszka 2024c). Badań przekrojowych, zarówno krajowych, jak i międzynarodowych, nie uzupełniano badaniami podłużnymi: naukowcy należący do klas osiagających najwyższą produktywność nie byli szczegółowo analizowani w ujęciu czasowym, ponieważ nie dysponowano danymi z wielu lat na mikropoziomie poszczególnych naukowców. Ustrukturyzowane Big Data o dużej jakości, czystości i globalnym zasięgu otworzyły nowe możliwości, z których tu korzystamy (natomiast duże badania odniesione do 38 krajów OECD, w tym do Polski, prezentujemy w Kwiek i Szymula 2023; Kwiek i Szymula 2024).

Zadajemy zatem proste pytanie: czy udział najbardziej produktywnych naukowców w krajowej produkcji publikacyjnej zmieniał się w ciągu ostatnich 30 lat? Najbardziej produktywni naukowcy są tu pogrupowani w trzy klasy pod względem produktywności publikacyjnej i odpowiednio szeregowani w ramach każdej badanej dyscypliny STEMM.

Przeprowadziliśmy analizę ich roli w podziale na płeć, wiek akademicki (czyli doświadczenie publikacyjne – okres, jaki upłynął od ukazania się pierwszej publikacji), dyscypliny naukowe, typy instytucji (10 instytucji intensywnie prowadzących badania i wybranych do programu IDUB i wszystkie pozostałe) i czas. Wykorzystaliśmy surowe,

jednostkowe dane dotyczące publikacji i cytowań pochodzące z bazy Scopus, zarówno dane wyodrębnione dla wszystkich polskich autorów i instytucji, jak i dane udostępnione nam w ramach wieloletniej umowy z International Center for the Study of Research (ICSR) Lab i przetworzone przez nas na potrzeby tego badania.

Dorobek na poziomie indywidualnym, czyli każdego naukowca osobno (ograniczony do dwóch typów publikacji: artykuły zamieszczone w czasopiśmie i artykuły zamieszczone w materiałach konferencyjnych) został obliczony oddzielnie dla sześcioletnich okresów, trzech klas produktywności i 15 dyscyplin STEM. Porównywaliśmy dorobek górnych 10% naukowców pod względem produktywności i dorobek pozostałych 90% naukowców, górnych 5% i pozostałych 95% oraz górnego 1% i pozostałych 99%, poszukując stabilnych wzorców.

2. Dane i podejście metodologiczne

2.1. Dane

Wykorzystaliśmy dane o autorach i publikacjach pobrane między listopadem 2022 r. a styczniem 2023 r. z bazy danych Scopus (1 TB) dotyczące okresu 1992–2021. Sporządziliśmy listę wszystkich polskich instytucji aktywnych badawczo (łącznie 343, wszystkie typy instytucji, wszystkie sektory, bez progu minimalnej liczby publikacji rocznie) i sporządziliśmy pełne listy autorów publikacji dla każdej instytucji posiadającej identyfikator instytucjonalny w bazie Scopus. Przypisaliśmy afiliacje instytucjonalne wszystkim naukowcom oddzielnie w każdym sześcioletnim okresie w oparciu o ich pełne dane publikacyjne z każdego okresu, aby nie pominąć (występującego w niewielkiej skali) wymiaru mobilności między instytucjami.

Informacje o autorach obejmowały ich imiona i nazwiska, dyscypliny, w których najczęściej publikowali swoje artykuły (wartość modalną dyscyplin wyliczoną ze wsparciem chmury dostarczonej przez ICSR Lab), afiliacje oraz unikalny identyfikator Scopus autora. Ponadto, korzystając z surowej bazy danych ICSR Lab, uzyskaliśmy rok pierwszej publikacji każdego naukowca, który wykorzystaliśmy do zdefiniowania wieku akademickiego w momencie opublikowania dowolnego artykułu i na koniec każdego sześcioletniego okresu.

Dyscypliny inne niż STEM oraz dyscypliny o małej liczbie autorów w naszej próbie (DEC, HEALTH, IMMU, NURS, VET, BUS, ECON, HUM, PSYCH, SOC, DENT i MULTI, korzystając z klasyfikacji ASJC, All Science Journal Classification) zostały usunięte z dalszej analizy. Dla niektórych dyscyplin sensowny binarny podział naszej próby na górnych 10%, 5% i 1% oraz resztę naukowców pod względem produktywności był niemożliwy: rozkład produktywności był zbyt płaski. Dopiero na tym etapie wstępnego przetwarzania danych zbiorów metadanych artykułów został powiązany determinis-

tycznie (przy użyciu indywidualnych identyfikatorów autorów Scopus jako klucza) z bazą danych autorów publikacji z polską afiliacją. Ostatecznie wykorzystany zbiór danych zawierał $N=152\,043$ polskich autorów z 15 dyscyplin (tabela 1) i ich $N_{\text{art}}=587\,558$ artykułów z okresu 1992–2021.

Nasze podejście do wyboru subpopulacji najbardziej produktywnych naukowców było stosunkowo proste: uszeregowaliśmy pod względem produktywności publikacyjnej wszystkich naukowców (i ich górne warstwy 1%, 5% i 10%) oddzielnie w ramach sześcioletnich okresów i oddzielnie w ramach dyscyplin. W zależności od ich wieku akademickiego niektórzy naukowcy (i niektórzy najbardziej produktywni naukowcy) pojawili się w jednym okresie, a inni w kilku lub we wszystkich okresach. Aby naukowcy zostali uwzględnieni w obliczeniach dla danego okresu, musieli być autorami co najmniej jednego artykułu opublikowanego w bazie Scopus w tym okresie.

Sześcioletnie okresy (zastosowana segmentacja sześcioletnia) zostały wybrane po to, aby zagwarantować, że liczba najbardziej produktywnych naukowców w każdej dyscyplinie w podziale na mężczyzn i kobiety w każdym okresie będzie spełniać minimalne wymogi ilościowe (było to szczególnie ważne w przypadku dyscyplin ilościowo zdominowanych przez mężczyzn, takich jak COMP, ENG, MATH i PHYS). Zbadaliśmy jednak również segmentację czteroletnią i pięcioletnią i potwierdziliśmy, że uzyskane wyniki nie są wrażliwe na różne segmentacje czasowe.

Podzbiory, z których wybrano najbardziej produktywnych naukowców, są większe w każdym kolejnym okresie: ich wielkość rośnie z 23 166 naukowców w okresie 1992–1997 do 93 092 w okresie 2016–2021 (tabela 1). Automatycznie zatem coraz większe są klasy górnych 10%, 5% i 1% naukowców pod względem produktywności: na przykład klasa górnych 10% rośnie z 2400 naukowców w pierwszym okresie do 9337 naukowców w ostatnim okresie (tabela 2).

Podzbiory naukowców, spośród których wybrano najbardziej produktywnych naukowców, mogą nakładać się na siebie w pięciu okresach: niektórzy naukowcy publikowali w pierwszych dwóch okresach (1992–1997 i 1998–2003), a inni w drugim i trzecim okresie (1998–2003 i 2004–2009); wreszcie najmłodszy naukowcy publikowali tylko w ostatnim okresie (2016–2021). Nasi unikalni autorzy (152 043 naukowców) pojawiają się w dowolnym okresie lub we wszystkich okresach; stąd liczba naukowców we wszystkich okresach (tabela 1) nie sumuje się do 152 043, ale jest wyższa.

Badaliśmy wszystkich naukowców publikujących w dowolnym okresie (lub we wszystkich okresach) między 1992 a 2021 rokiem. Z tego powodu ci sami naukowcy mogli być najbardziej produktywni (1%, 5%, 10%) w jednym okresie i należeć do reszty naukowców (99%, 95%, 90%) w innych okresach. Rozkład najbardziej produktywnych naukowców według klasy produktywności, okresu, płci, typu instytucji i grupy wieku akademickiego przedstawiono w tabeli uzupełniającej 1.

Tabela 1. Rozkład próby, z której wybrano najbardziej produktywnych naukowców według sześcioletniego okresu, płci, typu instytucji, grupy wieku akademickiego i dyscypliny STEM

Okres		1992 –1997	1998 –2003	2004 –2009	2010 –2015	2016 –2021
Łącznie		23 166	36 366	54 346	76 310	93 092
Płeć	Kobiety	8480	15 081	25 104	36 743	47 155
	Mężczyźni	14 686	21 285	29 242	39 567	45 937
Intensywność badawcza instytucji	IDUB	6338	9889	14 216	20 306	25 232
	Pozostałe	16 828	26 477	40 130	56 004	67 860
Grupa wieku akademickiego	0–9 lat	14 529	24 170	35 278	47 396	54 040
	10–19 lat	5297	6382	10 684	17 728	22 489
	20–29 lat	2853	4350	5132	6098	10 256
	30 lat i więcej	487	1464	3252	5088	6307
Dyscyplina	AGRI	1715	2916	4789	7649	9336
	BIO	2524	3766	5199	7280	8922
	CHEM	3318	4818	6159	7731	8221
	CHEMENG	220	375	491	535	570
	COMP	188	338	762	1620	1831
	EARTH	951	1594	1971	2712	3367
	ENER	50	109	223	559	1036
	ENG	1 065	1788	3370	5681	7952
	ENVIR	421	767	1456	2636	3928
	MATER	891	1331	2207	3730	5019
	MATH	898	1248	1705	2204	2244
	MED	6900	12 224	20 143	26 929	33 167
	NEURO	230	366	455	652	892
	PHARM	375	410	420	509	535
PHYS	3420	4316	4996	5883	6072	

Tabela 2. Rozkład podpróby najbardziej produktywnych naukowców (10%, 5% i 1% pod względem produktywności) według sześcioletniego okresu

	Górnych 10%	Górnych 5%	Górny 1%
1992–1997	2400	1202	241
1998–2003	3702	1846	374
2004–2009	5463	2726	550
2010–2015	7682	3841	770
2016–2021	9337	4666	934

Jedna czwarta wszystkich naukowców w pięciu badanych okresach pochodzi z uczelni intensywnie prowadzących badania (czyli w naszym uproszczeniu – z 10 uczelni objętych programem IDUB), a pozostałe trzy czwarte z pozostałych instytucji. Odsetek najbardziej produktywnych naukowców pochodzących z uczelni intensywnie prowadzących badania pozostaje prawie niezmienny w analizowanym okresie (i znajduje się w przedziale 25–32%, i jest najwyższy w przypadku górnego 1%; tabela uzupełniająca 1, środkowy panel).

Odsetek ten jest nadspodziewanie niski (a mamy na tej liście m.in. UW, UJ, UAM, AGH czy PW, tradycyjnie największe polskie uczelnie), stabilny w czasie, a ponadto tylko nieznacznie rośnie w miarę przesuwania się w górę klas wysokiej produktywności, od górnych 10% do górnego 1%. Naukowcy pochodzący z uczelni intensywnie prowadzących badania nie są wcale nadreprezentowani, jak można by się spodziewać: ich udział w subpopulacji najbardziej produktywnych naukowców niemal dokładnie odzwierciedla ich udział w całej populacji naukowców.

Skupiliśmy się na 15 dyscyplinach z obszaru STEMM, zgodnie z systemem klasyfikacji czasopism stosowanym w bazie Scopus (ASJC): AGRI, nauki rolnicze i biologiczne; BIO, biochemia, genetyka i biologia molekularna; CHEMENG, inżynieria chemiczna; CHEM, chemia; COMP, informatyka; EARTH, nauki o Ziemi i planetach; ENER, energia; ENG, inżynieria; ENVIR, nauka o środowisku; MATER, inżynieria materiałowa; MATH, matematyka; MED, nauki medyczne; NEURO, neuronauka; PHARM, farmakologia, toksykologia i farmacja; oraz PHYS, fizyka i astronomia.

2.2. Podejście metodologiczne

Zastosowaliśmy następujące podejście metodologiczne:

- 1) *Podłużne (longitudinalne)*: obserwujemy naukowców w czasie w ramach sześcioletnich okresów, maksymalnie w latach 1992–2021.
- 2) *Relatywne*: porównujemy naukowców, wykorzystując rangi percentylowe ich produktywności publikacyjnej (przypisując ich do odpowiednich średnich percentyli produktywności, od najwyższych do najniższych).
- 3) *Klasyfikujące*: klasyfikujemy naukowców na podstawie rangowania ich produktywności i porównujemy wyniki dla trzech klas najbardziej produktywnych naukowców z wynikami dla pozostałych naukowców.
- 4) *Znormalizowane do prestiżu czasopisma*: uznajemy za istotną przeciętną różnicę między publikowaniem w czasopismach o ogólnie niskim wpływie na globalny obieg naukowy i ogólnie wysokim wpływie na ten obieg.

Naszą jednostką analizy jest pojedynczy naukowiec z indywidualną historią publikacyjną w dowolnym sześcioletnim okresie. Oddzielnie w ramach 15 dyscyplin STEMM uszeregowaliśmy wszystkich naukowców widocznych na arenie międzynarodowej (tj.

naukowców publikujących artykuły badawcze w czasopismach indeksowanych w bazie Scopus) od najbardziej produktywnych do najmniej produktywnych i stworzyliśmy trzy klasy najbardziej produktywnych naukowców (górny 1%, 5% i 10%) oddzielnie w każdym analizowanym okresie.

Każdemu naukowcowi przypisaliśmy dominującą dyscyplinę akademicką w każdym okresie. Dominujące dyscypliny zostały obliczone na podstawie wartości modalnej (lub losowo wybranej wartości spośród wartości dominujących, jeśli było ich więcej niż jedna) wszystkich dyscyplin ASJC przypisanych do czasopism wszystkich cytowanych odniesień (*references, bibliography*) we wszystkich artykułach w publikacjach każdego naukowca z danego okresu.

Innymi słowy, korzystając z surowej bazy danych ICSR Lab i rozwiązań chmurowych, wszystkie cytowane odniesienia zawarte w bibliografiach wszystkich publikacji danego autora z danego okresu zostały powiązane z dyscyplinami ASJC i wybrano ich wartość modalną dla danego okresu.

Naukowcy mogli zatem zostać przypisani do tych samych lub różnych dyscyplin STEMM w miarę rozwoju ich kariery – co jest bardziej szczegółowym podejściem do przypisywania dyscyplin niż wykorzystywanie wszystkich cytowanych odniesień ze wszystkich publikacji (z całego życia) w celu wybrania jednej dominującej dyscypliny dla całej kariery, tak jak to robiliśmy w badaniach skoncentrowanych na 38 krajach OECD (Kwiek i Szymula 2023; Kwiek i Szymula 2024).

Użyliśmy dwucyfrowych (27 dyscyplin) kodów klasyfikacji Scopus, a nie kodów czterocyfrowych (333 obszary), ponieważ liczba obserwacji na bardziej szczegółowym poziomie była zbyt ograniczona, aby udało się przeprowadzić sensowną analizę, zwłaszcza dla najstarszego okresu (1992–1997) i dla analiz reprezentacji kobiet wśród najbardziej produktywnych naukowców w dyscyplinach zdominowanych przez mężczyzn. Płeć naukowców została zdefiniowana przy użyciu oprogramowania do określania płci *genderize.io*, z progiem prawdopodobieństwa ustawionym na poziomie 0,85, który uznano za zwracający najbardziej adekwatne wyniki.

Nasze podejście do produktywności publikacyjnej obejmuje normalizację do prestiżu czasopisma (i wykorzystuje metodę pełnego zliczania, w ramach której każdy współautor w publikacjach wieloautorskich otrzymuje równy udział w ich autorstwie). Nasza metoda (użyta już w Kwiek i Roszka 2024a) podkreśla różnicę przeciętnych wysiłków naukowych i przeciętnego wpływu na globalną społeczność naukową poprzez cytowania między artykułami opublikowanymi w ogólnie nieselektywnych czasopismach o niskim prestiżu oraz w ogólnie wysoce selektywnych czasopismach o wysokim prestiżu (patrz badanie tych samych artykułów – duplikatów – wydanych w różnych czasopismach i różnice w ich późniejszych cytowaniach w Larivière i Gingras 2010). Chociaż praca z metrykami wpływu opartymi na artykułach byłaby bardziej odpowiednia, ograniczenia na-

szego zbioru danych zmusiły nas do korzystania wyłącznie z metryk wpływu opartych na czasopismach.

Zastosowana normalizacja do prestiżu czasopisma opiera się na cytowaniach, a rangi percentylowe czasopism są określane corocznie w bazie Scopus na podstawie wszystkich cytowań otrzymanych przez wszystkie dokumenty opublikowane przez dane czasopismo w stosunku do ich liczby (w ciągu ostatnich czterech lat). W rezultacie artykułom zamieszczonym w czasopismach znajdujących się wyżej w rangach percentylowych Scopus CiteScore ($N=43\,092$ czasopisma, 2024) przypisano większą wagę w obliczaniu produktywności niż artykułom opublikowanym w czasopismach znajdujących się niżej (zakres: 0–99; wartościom 10 i mniejszym przyporządkowaliśmy 10). Dla całego okresu analizy stosowaliśmy rangi dla 2023 r., rangi historyczne nie są dostępne.

W nieznormalizowanym podejściu do produktywności (metoda pełnego zliczania) artykuły opublikowane we wszystkich czasopismach otrzymałyby wartość 1, podczas gdy w naszym znormalizowanym do prestiżu podejściu do produktywności (metoda pełnego zliczania) artykuły w czasopismach o randze percentylowej 90 otrzymywałyby wartość 0,90; artykuły opublikowane w czasopismach o randze percentylowej 10 i niższej otrzymywałyby wartość 0,1 (zob. zmiany klas produktywności naukowców w wybranych okresach dla 2326 polskich profesorów tytularnych w Kwiek i Roszka 2024a i dla 4165 polskich doktorów habilitowanych w Kwiek i Roszka 2024b).

3. Wyniki

3.1. Wzorce empiryczne: udział najbardziej produktywnych naukowców w krajowej produkcji naukowej na przestrzeni 30 lat (1992–2021)

Najbardziej produktywni naukowcy mają znaczący – i niemal niezmienny w czasie – udział w krajowej produkcji naukowej w Polsce w pięciu 6-letnich okresach, na które podzieliliśmy ostatnich 30 lat. W przypadku górnych 10% odsetek ten w ostatnim badanym okresie wyniósł 45,5% dla wszystkich dyscyplin łącznie, a w przypadku górnego 1% – wyniósł 10,8%. Odsetki dla wszystkich okresów mieszczą się w przedziale 43,8–46,6% dla górnych 10% naukowców i 10,1–11,6% dla górnego 1%, z interesującymi różnicami zachodzącymi między dyscyplinami (tabela 3).

Nasze badania pokazują, że z perspektywy długoterminowej struktura polskiej produkcji naukowej pod względem publikacji najbardziej produktywnych naukowców w porównaniu z publikacjami reszty naukowców pozostaje zaskakująco stabilna.

W najprostszym ujęciu: średnio 1% naukowców odpowiada za 10% krajowej produkcji publikacyjnej, a 10% naukowców odpowiada za prawie połowę wszystkich publikacji. Proporcje te praktycznie nie zmieniają się w czasie, co oznacza, że nie zachodzą procesy rosnącej koncentracji badań w rękach najbardziej produktywnych naukowców. W każdym z analizowanych okresów obserwuje się znaczne zróżnicowanie dyscyplinarne.

Tabela 3. Publikacje 1% i 10% najbardziej produktywnych naukowców jako odsetek wszystkich publikacji (produkcja krajowa) według dyscypliny STEMM i okresu (w %)

Górny 1%					
Dyscyplina / okres	1992 –1997	1998 –2003	2004 –2009	2010 –2015	2016 –2021
AGRI	9,0	10,0	9,4	10,0	9,3
BIO	9,6	9,3	9,4	9,6	8,9
CHEM	9,7	10,8	11,9	12,1	11,3
CHEMENG	15,5	13,8	11,2	11,8	9,7
COMP	8,4	9,8	11,9	12,5	12,0
EARTH	9,2	10,5	10,9	10,1	9,2
ENER	7,6	9,4	8,4	13,8	11,0
ENG	9,2	11,2	12,0	11,2	11,1
ENVI	11,1	13,8	13,4	10,7	10,1
MATER	9,5	10,0	10,6	9,8	10,3
MATH	7,9	7,9	8,2	8,7	8,9
MED	10,7	10,8	12,9	12,4	11,9
NEURO	8,4	7,4	9,6	9,1	8,7
PHARM	19,2	15,6	9,4	12,0	13,2
PHYS	10,5	11,9	11,8	12,9	11,6
ŁĄCZNIE	10,1	10,7	11,6	11,5	10,8
Górnych 10%					
Dyscyplina / okres	1992 –1997	1998 –2003	2004 –2009	2010 –2015	2016 –2021
AGRI	39,0	42,5	42,2	43,1	42,7
BIO	41,7	41,8	41,7	43,0	41,8
CHEM	45,0	46,5	47,7	47,3	45,6
CHEMENG	46,5	47,3	46,1	44,9	45,6
COMP	39,5	45,9	44,6	45,9	45,0
EARTH	43,8	45,0	45,1	43,9	41,3
ENER	33,2	35,2	38,0	46,7	43,6
ENG	41,4	44,1	44,2	44,5	42,3
ENVI	44,1	46,1	46,2	43,6	43,5
MATER	42,5	44,6	44,2	43,9	45,8
MATH	35,6	37,8	40,0	41,6	39,3
MED	45,7	46,6	49,8	49,1	49,0
NEURO	41,7	42,2	47,2	41,3	39,2
PHARM	62,1	58,5	52,3	52,7	58,1
PHYS	44,1	46,8	45,8	47,2	45,1
ŁĄCZNIE	43,8	45,4	46,6	46,4	45,5

Jednak najbardziej produktywni naukowcy nie tworzą jednorodnych klas. Są reprezentowani przez mężczyzn i kobiety, są powiązani z instytucjami bardziej lub mniej intensywnie prowadzącymi badania (IDUB vs. pozostałe); i wreszcie należą do różnych grup wieku akademickiego (grup doświadczenia publikacyjnego obliczanego na podstawie czasu, jaki upłynął od ich pierwszej publikacji) (tabela 4). Używając wieku akademickiego jako zamiennika wieku biologicznego (z korelacjami analizowanymi szczegółowo dla Polski dla dyscyplin STEM na poziomie 0,8–0,9 w Kwiek i Roszka 2022b), określamy ich jako naukowców początkujących, naukowców na średnim etapie kariery, dojrzałym etapie kariery i późnym jej etapie (odpowiednio grupy 0–9, 10–19, 20–29 oraz 30 i więcej lat doświadczenia publikacyjnego).

Tabela 4. Publikacje 10% najbardziej produktywnych naukowców jako odsetek wszystkich publikacji (produkcja krajowa) według płci, typu instytucji i grupy wiekowej, z uwzględnieniem wszystkich dyscyplin STEM łącznie (w %)

Okres	Łącznie	Płeć		Intensywność badawcza instytucji		Grupa wieku akademickiego (długość doświadczenia publikacyjnego)			
		Kobiety	Mężczyźni	IDUB	Pozostałe	0–9 lat	10–19 lat	20–29 lat	30 i więcej lat
1992–1997	43,8	34,7	47,6	41,7	44,7	26,3	50,5	62,0	68,0
1998–2003	45,4	36,0	50,1	45,8	45,2	25,3	55,5	61,6	66,9
2004–2009	46,6	37,3	52,3	47,4	46,3	24,1	56,2	62,6	69,6
2010–2015	46,4	37,8	52,4	47,7	45,9	22,6	54,4	65,0	67,8
2016–2021	45,5	38,4	51,0	46,0	45,2	23,0	52,5	61,6	61,7

Nasze analizy pokazują znaczne różnice między kobietami i mężczyznami w grupie 10% najbardziej produktywnych naukowców (tabela 4). Kobiety znajdujące się w górnych 10% odpowiadają za około jedną trzecią wszystkich publikacji wydanych przez kobiety (tj. pozostałe kobiety odpowiadają za około dwie trzecie wszystkich publikacji wydanych przez kobiety) i udział ten rośnie z czasem. Natomiast koncentracja badań jest znacznie wyższa wśród mężczyzn z górnych 10%.

Mężczyźni z górnych 10% są niezmiennie odpowiedzialni za ponad 50% wszystkich publikacji wydawanych przez mężczyzn. Różnice między 10% najbardziej produktywnych naukowców powiązanych z uczelniami z dzisiejszej listy IDUB a wszystkimi innymi instytucjami są marginalne w ciągu badanych 30 lat. We wszystkich okresach wśród nau-

kowców na średnim i późnym etapie kariery naukowej udział publikacji autorstwa najbardziej produktywnych naukowców przekracza 60% – im starsi naukowcy, tym większa jest koncentracja publikacji w ich rękach (z wielu powodów, między innymi takich jak kierowanie dużymi zespołami badawczymi).

3.2. Wskaźnik Względnej Obecności (WVO) w klasie najbardziej produktywnych naukowców dla mężczyzn i kobiet

W tej części pracy interesuje nas względna obecność kobiet wśród najbardziej produktywnych naukowców. Normalizujemy liczbę kobiet wśród nich w danym okresie i w danej dyscyplinie do liczby wszystkich kobiet w populacji naukowców w tym okresie i w tej dyscyplinie.

Skonstruowaliśmy Wskaźnik Względnej Obecności (WVO) wśród najbardziej produktywnych naukowców dla mężczyzn i dla kobiet. Wskaźnik WVO równy 1 dla kobiet oznacza taką samą względną obecność mężczyzn i kobiet w tej klasie. Wskaźnik ten dla kobiet może pokazywać poziom ich nadreprezentacji (wartość powyżej 1) lub niedoreprezentacji (wartość poniżej 1) wśród najbardziej produktywnych naukowców. Chociaż liczby (w podejściu nominalnym) mogą potwierdzać prosty fakt, że najbardziej produktywni mężczyźni w danej dyscyplinie przewyższają kobiety liczebnie, wskaźnik WVO dla kobiet zapewnia bardziej adekwatną i intuicyjnie zrozumiałą miarę ich obecności w tej klasie.

Wskaźnik ten dla mężczyzn jest konstruowany jako liczba najbardziej produktywnych mężczyzn w stosunku do całkowitej liczby mężczyzn (w danej domenie) podzielona przez liczbę najbardziej produktywnych kobiet w stosunku do całkowitej liczby kobiet (w danej domenie).

Wskaźnik może zmieniać się w czasie i różnić się w zależności od dyscypliny. Z technicznego punktu widzenia wskaźnik ten jest ilorazem dwóch ilorazów: mężczyźni najbardziej produktywni /wszyscy mężczyźni podzieleni przez najbardziej produktywne kobiety/ wszystkie kobiety (zob. Abramo et al. 2009 na temat „gwiazd nauki”). Zależność między wskaźnikiem dla mężczyzn i wskaźnikiem dla kobiet jest harmoniczna (a nie liniowa). Jeśli kobiety w danej domenie wykazują wskaźnik na poziomie 0,33, to mężczyźni wykazują go na poziomie 3,03 (1/0,33) w tej domenie; jeśli kobiety w danej domenie wykazują wskaźnik na poziomie 0,5, to mężczyźni wykazują w niej ten wskaźnik na poziomie 2 (1/0,5) itd.

Jak należy interpretować wskaźniki 1,29 dla mężczyzn i 0,77 dla kobiet dla AGRI dla 10% najbardziej produktywnych naukowców (w latach 1992–1997, tabela 5, górny panel)? Wskaźniki te pokazują, że reprezentacja mężczyzn w górnych 10% najbardziej produktywnych naukowców jest o 29% wyższa niż reprezentacja kobiet w górnych 10% w tej dyscyplinie w tym okresie. Mężczyźni są nadreprezentowani.

Tabela 5. Wskaźnik względnej obecności (WWO) wśród najbardziej produktywnych naukowców dla mężczyzn (lewe panele) i kobiet (prawe panele), górnych 10% i 1% (pod względem produktywności) w ujęciu sześciolletnich okresów, grup wieku akademickiego i dyscypliny STEM

Zmienna	Kategoria / okres	Mężczyźni					Kobiety				
		1992-1997	1998-2003	2004-2009	2010-2015	2016-2021	1992-1997	1998-2003	2004-2009	2010-2015	2016-2021
Górnych 10%											
Grupa wieku akademickiego	0-9 lat	1,63	1,40	1,50	1,48	1,73	0,61	0,71	0,66	0,68	0,58
	10-19 lat	1,29	1,31	1,31	1,25	1,19	0,77	0,76	0,76	0,80	0,84
	20-29 lat	1,37	1,20	1,18	1,16	1,08	0,73	0,83	0,85	0,86	0,92
	30 lat i więcej	1,03	1,28	1,16	1,15	1,06	0,97	0,78	0,86	0,87	0,94
Dyscyplina	AGRI	1,29	1,42	1,56	1,46	1,52	0,77	0,70	0,64	0,69	0,66
	BIO	2,19	2,33	2,51	2,08	1,91	0,46	0,43	0,40	0,48	0,52
	CHEM	1,97	1,96	1,98	1,98	1,83	0,51	0,51	0,50	0,51	0,55
	CHEMENG	1,91	1,75	1,86	2,17	2,35	0,52	0,57	0,54	0,46	0,42
	COMP	1,55	0,63	0,88	0,94	2,00	0,65	1,59	1,14	1,06	0,50
	EARTH	2,05	1,86	1,96	1,89	1,35	0,49	0,54	0,51	0,53	0,74
	ENER	0,44	1,19	1,69	1,07	1,76	2,28	0,84	0,59	0,94	0,57
	ENG	3,03	1,95	1,86	1,97	1,74	0,33	0,51	0,54	0,51	0,58
	ENVIR	1,01	0,99	1,04	1,29	1,11	0,99	1,01	0,96	0,78	0,90
	MATER	1,37	1,51	1,53	1,41	1,43	0,73	0,66	0,65	0,71	0,70
MATH	3,35	18,14	2,71	1,88	1,14	0,30	0,06	0,37	0,53	0,88	
MED	1,34	1,46	1,75	1,84	1,79	0,75	0,68	0,57	0,54	0,56	
NEURO	1,21	1,46	1,22	1,37	1,00	0,82	0,68	0,82	0,73	1,00	
PHARM	0,98	0,74	0,89	1,26	1,61	1,02	1,35	1,13	0,80	0,62	

	PHYS	1,90	1,66	1,82	1,57	1,60	0,53	0,60	0,55	0,64	0,63	
	ŁĄCZNIE	1,56	1,60	1,71	1,66	1,59	0,64	0,62	0,58	0,60	0,63	
Górny 1%												
Grupa wieku akademickiego	0-9 lat	2,94	2,82	3,54	2,54	3,75	0,34	0,35	0,28	0,39	0,27	
	10-19 lat	3,55	1,79	2,14	1,96	2,16	0,28	0,56	0,47	0,51	0,46	
	20-29 lat	3,16	2,26	1,57	1,66	1,80	0,32	0,44	0,64	0,60	0,55	
	30 lat i więcej	2,45	1,98	1,57	1,70	1,72	0,41	0,51	0,64	0,59	0,58	
	AGRI	4,28	2,57	2,75	3,63	2,73	0,23	0,39	0,36	0,28	0,37	
	BIO	4,63	6,32	8,76	5,42	3,84	0,22	0,16	0,11	0,11	0,18	0,26
	CHEM	9,01	3,60	2,74	2,64	2,61	0,11	0,28	0,36	0,38	0,38	
	CHEMENG	-	-	-	-	-	-	-	-	-	-	-
	COMP	-	-	1,27	0,79	3,71	-	-	-	0,79	1,26	0,27
	EARTH	-	3,32	2,17	1,96	2,09	-	-	0,30	0,46	0,51	0,48
Dyscyplina	ENER	-	0,24	-	-	2,09	-	4,19	-	-	0,48	
	ENG	-	-	5,90	5,04	2,59	-	-	0,17	0,20	0,39	
	ENVIR	-	2,98	2,17	2,09	1,90	-	0,34	0,46	0,48	0,53	
	MATER	0,51	3,11	2,46	1,95	2,72	1,95	0,32	0,41	0,51	0,37	
	MATH	-	-	5,79	1,11	3,25	-	-	0,17	0,90	0,31	
	MED	2,65	2,60	2,84	3,35	4,05	0,38	0,38	0,35	0,30	0,25	
	NEURO	1,32	3,08	2,69	1,27	1,55	0,76	0,32	0,37	0,79	0,65	
	PHARM	4,11	0,36	1,71	0,51	1,61	0,24	2,80	0,58	1,95	0,62	
	PHYS	8,89	2,20	2,77	3,99	3,56	0,11	0,45	0,36	0,25	0,28	
		ŁĄCZNIE	3,64	2,92	2,95	2,85	2,97	0,27	0,34	0,34	0,35	0,34

Uwaga: „-” oznacza brak kobiet naukowców w danej kategorii

I odwrotnie: reprezentacja kobiet w górnych 10% stanowi tylko 77% reprezentacji mężczyzn w górnych 10% w tej dyscyplinie w tym okresie. Kobiety są niedoreprezentowane (jednocześnie związek między dwoma wskaźnikami jest harmoniczny: $1/1,29 = 0,77$). Aby pokazać dwie perspektywy (czyli ogólnie nadreprezentację mężczyzn i niedoreprezentację kobiet), pozostawiamy wskaźniki zarówno dla mężczyzn, jak i dla kobiet w tabeli 5, chociaż łączy je prosta relacja.

Względna nadreprezentację mężczyzn wśród 10% i 1% najbardziej produktywnych naukowców (tabela 5) można analizować w kilku wymiarach: według okresu, czyli pokazując zmiany w czasie; według grupy wieku akademickiego, czyli pokazując zmiany według doświadczenia publikacyjnego (niezwykle silnie skorelowanego z wiekiem biologicznym, zob. Kwiek i Roszka 2022a); i według dyscypliny.

Ogólne wzorce pokazują, że mężczyźni są nadreprezentowani w każdej klasie najbardziej produktywnych naukowców (1%, 5% i 10%), a nadreprezentacja ta rośnie wraz z przesuwaniami się w górę skali wysokiej produktywności. W ostatnim badanym okresie (2016–2021) wskaźnik WWO dla 10% dla mężczyzn wynosi 1,59, a dla 1% dla mężczyzn wynosi aż 2,97 (wszystkie dyscypliny łącznie: tabela 5). Wskaźnik maleje dla naukowców z długim doświadczeniem akademickim, a dla 10% zbliża się do 1,00, choć nigdy go nie osiąga. Wskaźnik dla mężczyzn rośnie wraz z przesuwaniami się w górę skali wysokiej produktywności, z bardzo dużą nadreprezentacją mężczyzn wśród 1% najbardziej produktywnych naukowców. W pięciu badanych okresach, dla górnych 10% i 5%, wskaźnik pozostaje na podobnym poziomie. Ponadto wskaźnik ten jest silnie zróżnicowany pod kątem dyscyplin. Wydaje się jednak, że nie zachodzi prosta zależność między wskaźnikiem a strukturą dyscyplin w ujęciu płci.

3.3. Podejście wielowymiarowej regresji logistycznej

3.3.1. Logitowe uogólnione modele liniowe z efektami stałymi

W celu detekcji determinant przynależności do klas najbardziej produktywnych naukowców opracowaliśmy modele ekonometryczne. Ze względu na binarny charakter zmiennej zależnej wykorzystaliśmy model regresji logistycznej z efektami stałymi (Allison 2000). Model ten ma charakter przekrojowy i dynamiczny (Fernández-Val i Weidner 2016; Hinz i in. 2020; Stammann 2018), a efektami stałymi w modelu były okresy i dyscypliny STEM. Wykorzystano pięć sześcioletnich okresów publikacyjnych (w latach 1992–2021) i 15 dyscyplin STEM. Dyscypliny w badanych okresach były dominującymi (wartości modalne) w pełnych listach cytowanych odniesień autorów we wszystkich ich publikacjach z danego okresu, zgodnie z klasyfikacją czasopism Scopus ASJC opartą na 27 dyscyplinach (z dwucyfrowymi kodami). Zmienne niezależne i zmienne efektów stałych opisano w tabeli 6.

Tabela 6. Opis zmiennych niezależnych i zmiennych efektów stałych

Zmienna	Opis	Typ
Wiek akademicki w okresie	Średnia różnica między rokiem kończącym okres (1997, 2003, 2009 etc.) i rokiem pierwszej publikacji autora	Ilościowa ciągła
Średnia wielkość zespołu w okresie	Średnia arytmetyczna wielkości zespołu (średnia liczby współautorów plus 1 dla wszystkich artykułów) w danym okresie	Ilościowa ciągła
Wskaźnik współpracy ogólnej w okresie	Odsetek artykułów napisanych we współpracy: stosunek liczby artykułów z co najmniej jednym współautorem do całkowitej liczby artykułów w danym okresie (zakres 0–100%).	Ilościowa ciągła
Wskaźnik współpracy międzynarodowej w okresie	Odsetek artykułów napisanych we współpracy międzynarodowej: stosunek liczby artykułów napisanych we współpracy z autorem z afiliacją zagraniczną do całkowitej liczby artykułów w danym okresie (zakres 0–100%).	Ilościowa ciągła
Płeć	Płeć autora (według <i>genderize.io</i> , próg prawdopodobieństwa 0,85)	Jakościowa binarna (mężczyzna, kobieta)
Intensywność badawcza instytucji	Afiliacja instytucjonalna naukowca z jedną z 10 uczelni uznanych arbitralnie za instytucje intensywnie prowadzące badania (udział w pełnym 10% programie IDUB w latach 2020–2025)	Jakościowa binarna (IDUB, pozostałe)
Dyscyplina w danym okresie	Dominująca dyscyplina (wartość modalna) wyliczona z pełnych list cytowanych przez autora odniesień we wszystkich publikacjach z danego okresu, zgodnie z klasyfikacją czasopism Scopus ASJC, 15 pól tematycznych, zastosowano kody 2-cyfrowe	Jakościowa, 15-wariantowa
Okres	Sześcioletni okres publikacyjny, pięć okresów w latach 1992–2021	Jakościowa, 5-wariantowa

Model ma charakter przekrojowy, ponieważ populacja składa się z indywidualnych naukowców (jednostką analizy jest pojedynczy naukowiec opisany wektorem cech), i dynamiczny, ponieważ obejmuje pięć rozłącznych sześcioletnich okresów. Włączenie zmiennej czasowej do modelu jako zmiennej niezależnej zakłóciłoby założenie o niezależności obserwacji, ponieważ poszczególni naukowcy mogą publikować w więcej niż jednym okresie (a niektórzy publikują we wszystkich okresach). Jednocześnie poszczególne okresy różniły się między sobą warunkami społecznymi i ekonomicznymi (wpływającymi na funkcjonowanie publicznego systemu nauki).

Wybór dyscyplin jako drugiego efektu stałego był podyktowany wewnętrznymi różnicami we wzorcach publikowania w ramach dyscyplin oraz możliwością dokonywania porównań między dyscyplinami bez zniekształceń związanych z wpływem czynników specyficznych dla dyscypliny na ogólną produktywność. Do obliczeń wykorzystano oprogramowanie R w wersji 4.3.0, RStudio w wersji 2023.06.0 Build 421 oraz pakiet „alpaca” w wersji 0.3.4.

3.3.2. Zmienne niezależne

Dobór zmiennych (tabela 6) był podyktowany dostępnością danych i literaturą przedmiotu dotyczącą wysokiej produktywności badawczej (Kwiek 2016; Kwiek 2018a). Ze względu na panelowy charakter badania zastosowano model regresji logistycznej z efektami stałymi. Utworzono trzy modele dla trzech klas najbardziej produktywnych naukowców (1%, 5% i 10%; tabela 7). Aby sprawdzić założenie o braku współliniowości wektora zmiennych niezależnych, zastosowano metodę odwróconej macierzy korelacji i nie wykryto istotnego skorelowania wektora zmiennych niezależnych. Analizy przeprowadzono dla każdego okresu i dla każdej klasy najbardziej produktywnych naukowców.

3.3.3. Wyniki regresji logistycznej

W modelach regresji logistycznej analizujemy, które predyktory zmieniają prawdopodobieństwo wejścia do klas najbardziej produktywnych naukowców (tabela 7). Model dla górnych 10% najbardziej produktywnych naukowców identyfikuje kilka ważnych predyktorów, które zwiększają szanse na sukces: płeć, wiek akademicki i różne miary współpracy naukowej.

Mężczyźni mają znacznie większe szanse (średnio o 39,2%; $\text{Exp}(B) = 1,392$) w porównaniu z kobietami, z wąskim przedziałem ufności (1,354–1,432) podkreślającym wyraźny wpływ płci na wysoką produktywność publikacyjną. Każdy dodatkowy rok wieku akademickiego zwiększa szanse na sukces średnio o 7,2% ($\text{Exp}(B) = 1,072$; przedział ufności 1,071–1,073). Średnia wielkość zespołu, wskaźnik współpracy międzynarodowej i wskaźnik współpracy ogólnej – wyliczone na podstawie wszystkich publikacji każdego naukowca oddzielnie dla każdego okresu – również pozytywnie wpływają na prawdopodobieństwo bycia najbardziej produktywnym naukowcem. Każdy dodatkowy punkt procentowy wskaźnika współpracy ogólnej (w zakresie 0–100%) zwiększa prawdopodobieństwo sukcesu średnio o 0,5%. Wskaźnik współpracy międzynarodowej (również w zakresie 0–100%) ma większe znaczenie, a każdy dodatkowy punkt procentowy zwiększa szansę na sukces średnio o 1%, przy założeniu, że wszystkie inne czynniki pozostają niezmiennione (*ceteris paribus*). Ponadto wzrost średniej wielkości zespołu o jedną jednostkę (jednego współautora publikacji) zwiększa szansę na sukces średnio o 0,9%.

Tabela 7. Regresja logistyczna z oszacowaniem efektów stałych (efekty stałe: okresy i dyscypliny), oszacowania ilorazu szans przynależności do klasy górnych 10%, 5% i 1% najbardziej produktywnych naukowców

Klasa najbardziej produktywnych naukowców	Zmienna	Exp(B)	LB	UB	Pr (> z)
Górnych 10%	Wiek akademicki w okresie	1,072	1,071	1,073	< 0,001
	Średnia wielkość zespołu w okresie	1,009	1,007	1,011	< 0,001
	Wskaźnik współpracy międzynarodowej w okresie	1,010	1,009	1,010	< 0,001
	Wskaźnik współpracy ogólnej w okresie	1,005	1,004	1,005	< 0,001
	Płeć: mężczyzna	1,392	1,354	1,432	< 0,001
	Intensywność badawcza instytucji: pozostałe	0,953	0,925	0,981	0,001
Górnych 5%	Wiek akademicki w okresie	1,074	1,072	1,075	< 0,001
	Średnia wielkość zespołu w okresie	1,010	1,007	1,013	< 0,001
	Wskaźnik współpracy międzynarodowej w okresie	1,011	1,010	1,011	< 0,001
	Wskaźnik współpracy ogólnej w okresie	1,006	1,005	1,007	< 0,001
	Płeć: mężczyzna	1,597	1,536	1,661	< 0,001
	Intensywność badawcza instytucji: pozostałe	0,970	0,932	1,010	0,141
Górny 1%	Wiek akademicki w okresie	1,081	1,078	1,084	< 0,001
	Średnia wielkość zespołu w okresie	1,010	1,005	1,015	< 0,001
	Wskaźnik współpracy międzynarodowej w okresie	1,013	1,012	1,014	< 0,001
	Wskaźnik współpracy ogólnej w okresie	1,009	1,006	1,012	< 0,001
	Płeć: mężczyzna	2,252	2,055	2,469	< 0,001
	Intensywność badawcza instytucji: pozostałe	0,905	0,832	0,984	0,020

W przypadku 5% najbardziej produktywnych naukowców wpływ tych samych predyktorów jest jeszcze bardziej wyraźny. Bycie mężczyzną ($\text{Exp}(B) = 1,597$) i wiek akademicki ($\text{Exp}(B) = 1,074$) również w tym przypadku odgrywają znaczącą rolę, a wskaźnik współpracy międzynarodowej ($\text{Exp}(B) = 1,011$) i wskaźnik współpracy ogólnej wywierają znaczący wpływ na osiągnięcie sukcesu. Średnia wielkość zespołu ma nieco większy wpływ. Dla 1% najbardziej produktywnych naukowców efekty predyktorów osią-

gają swoje maksima. Płeć ($\text{Exp}(B) = 2,252$) i wiek akademicki ($\text{Exp}(B) = 1,081$) wykazują najsilniejszy wpływ. Znaczenie mają również wskaźnik współpracy międzynarodowej, wskaźnik współpracy ogólnej i średnia wielkość zespołu.

Co ciekawe, zatrudnienie poza 10 uczelniami uczestniczącymi dzisiaj w programie IDUB wykazuje jedynie niewielki negatywny związek z szansami wejścia do dwóch klas najbardziej produktywnych naukowców ($\text{Exp}(B) = 0,953$ dla 10% i $\text{Exp}(B) = 0,905$ dla 1%). Praca poza uczelniami z listy IDUB zmniejsza szansę na wejście do klasy najbardziej produktywnych naukowców jedynie o 5% dla klasy górnych 10% i o 10% dla klasy górnego 1% najbardziej produktywnych naukowców.

Siła wpływu wszystkich istotnych parametrów rośnie w miarę przesuwania się w górę skali produktywności, zwłaszcza rośnie wpływ płci i wieku akademickiego. Bycie mężczyzną jako predyktor wysokiej produktywności publikacyjnej znacznie zwiększa swój wpływ: $\text{Exp}(B)$ rośnie z 1,392 dla klasy 10% do 2,252 dla klasy 1%, a szanse mężczyzn są średnio większe o prawie 40% i aż o ponad 125%. Im bardziej selektywna jest klasa wysokiej produktywności, tym większe szanse na wejście do niej mają mężczyźni (a mniejsze szanse mają kobiety), przy założeniu stałego wpływu wszystkich pozostałych czynników.

Dla klasy 10% najbardziej produktywnych naukowców wartości efektów stałych dla każdego okresu maleją wraz z upływem czasu, co oznacza, że w kolejnych okresach coraz trudniej do niej należeć. Rośnie konkurencja ze względu na ogólny wzrost liczby publikacji i poszerzające się różnice między mniej i bardziej produktywnymi naukowcami. Najwyższe wartości dla takich dyscyplin, jak: ENG, MED i PHARM, sugerują, że konkurencja o przynależność do klasy 10% jest mniejsza (w porównaniu ze średnią dla modelu dla wszystkich dyscyplin). Natomiast najniższe wartości zaobserwowane dla CHEM, EARTH, MATH i PHYS wskazują na wysoką konkurencję w tych dyscyplinach (zob. tabela uzupełniająca 2). Modele wyjaśniają od 10,8% (górnym 10%) do 13,1% (górnym 1%) zmienności zmiennej zależnej.

Dla wszystkich klas najbardziej produktywnych naukowców wartości efektów stałych maleją wraz z zawężaniem się klas z 10% do 1%, co oznacza, że trudniej jest dostać się do wyższych klas. Zarazem dla każdej klasy najbardziej produktywnych naukowców wartości efektów stałych maleją z każdym kolejnym okresem, co oznacza, że konkurencja w dostaniu się do nich rośnie w czasie.

W każdym kolejnym okresie wejście do każdej klasy najbardziej produktywnych naukowców jest trudniejsze, przy czym w każdym przypadku najtrudniej jest wejść do klasy górnego 1% w ostatnim okresie (2016–2021). Tendencja spadkowa efektów stałych w kolejnych okresach wskazuje na rosnące wymagania dotyczące produktywności publikacyjnej pozwalające na znalezienie się wśród najbardziej produktywnych naukowców.

4. Podsumowanie

Uzyskane przez nas wyniki pokazują, że udział najbardziej produktywnych naukowców w krajowej produkcji publikacyjnej jest zaskakująco stabilny w czasie: chociaż świat społeczny i gospodarczy doświadczał potężnych transformacji, a nauka akademicka przechodziła prawdopodobnie największe zmiany w swojej najnowszej historii (podobnie jak całe szkolnictwo wyższe), najbardziej produktywni naukowcy w latach 1992–2021 zawsze odgrywali strukturalnie podobną, fundamentalną rolę.

Górny 1% naukowców odpowiadał średnio za 10% krajowej produkcji publikacyjnej (co można nazwać regułą 1/10), a górnych 10% naukowców odpowiadało za prawie 50% tej produkcji (co można nazwać regułą 10/50), przy znacznych różnicach dyscyplinarnych.

Przeprowadziliśmy analizę udziału naukowców z perspektywy czasowej (1992–2021) i dyscyplinarnej (15 dyscyplin STEMM). Trzy klasy najbardziej produktywnych naukowców (górny 1%, 5% i 10% naukowców pod względem produktywności publikacyjnej) zostały zbadane oddzielnie w ramach dyscyplin i sześcioletnich okresów.

Zbadaliśmy wszystkich polskich naukowców widocznych na arenie międzynarodowej – poprzez publikacje indeksowane w bazie Scopus (od $N=23\,166$ naukowców w pierwszym okresie do $N=93\,092$ naukowców w ostatnim okresie) oraz dużą próbę najbardziej produktywnych naukowców (od $N=2\,400$ w pierwszym okresie do $N=9\,337$ w ostatnim okresie dla górnych 10%).

Pod względem metodologicznym nasze badanie było podłużne (longitudinalne), relatywne, klasyfikujące i znormalizowane do prestiżu czasopisma. Jego podłużny charakter pozwolił na śledzenie naukowców w czasie przez cały analizowany okres, czyli przez trzy dekady lub tak długo, jak długo byli oni autorami publikacji (o rezygnacji z nauki w 38 krajach OECD, zob. Kwiek i Szymula 2024). Relatywny charakter badania pozwolił nam na szeregowanie naukowców według produktywności w ramach każdego okresu. Klasyfikujące podejście pozwoliło na zastosowanie binarnego rozróżnienia na trzy klasy najbardziej produktywnych naukowców i pozostałych naukowców (odpowiednio 99%, 95% i 90%). A jego charakter znormalizowany do prestiżu wykorzystał strukturę globalnego systemu czasopism naukowych (opartego na cytowaniach i dostarczanego przez bazę Scopus) do obliczeń indywidualnej produktywności.

Skalę potencjalnego udziału najbardziej produktywnych naukowców mogliśmy przewidzieć na podstawie naszych wcześniejszych badań przekrojowych, prowadzonych na małą skalę i opartych na 4000 zwróconych ankiet w Polsce i 17 000 w Europie. Jednak kompleksowe podejście podłużne, obejmujące trzy dekady, nie byłoby możliwe bez dostępu do pełnego zbioru surowych danych Scopus i rozwiązań chmurowych udostępnionych nam przez ICSR Lab.

Być może rozkład wysoce produktywnych naukowców wśród kolejnych pokoleń polskich naukowców jest względnie stabilny w czasie, ponieważ jest związany z procesami kumulacji przewag w nauce akademickiej na poziomie indywidualnym (co ma tym większe znaczenie im bardziej niedofinansowany jest system nauki). Może jest on także związany z prostym rozkładem indywidualnych motywacji, wrodzonych zdolności i wreszcie szczęścia w karierze naukowej (jak pokazuje ekonomia nauki, Stephan 2012; zob. szerzej o systemowych nierównościach w nauce w monografiach Kwiek 2015 i Kwiek 2019).

Populacja polskich naukowców może być strukturalnie podzielona na dwa stałe segmenty: pierwszy – mały i niezwykle produktywny (silnie zmotywowany, dobrze naukowo przygotowany, zdolny, skupiony na badaniach i dodatkowo mający szczęście) oraz drugi, wielokrotnie większy, obejmujący wszystkich pozostałych naukowców (mniej zmotywanych, mniej zdolnych, gorzej przygotowanych, mniej skupionych na badaniach, a być może również od początku kariery niemających szczęścia).

Klasy naukowców o najwyższej produktywności wyłaniają się z naszych badań jako klasy heterogeniczne pod względem płci i grup wieku akademickiego (czyli doświadczenia publikacyjnego). Kobiety w górnych 10% odpowiadają za około 35–37% wszystkich publikacji wydawanych przez kobiety. Koncentracja produkcji naukowej jest większa w przypadku mężczyzn w górnych 10%: mężczyźni są odpowiedzialni za ponad 50% wszystkich publikacji wydawanych przez mężczyzn.

Ponadto koncentracja produkcji rośnie wraz z grupami wieku akademickiego: najbardziej produktywni naukowcy na dojrzałym i późnym etapie kariery zawodowej – czyli naukowcy z co najmniej dwudziestoletnim doświadczeniem publikacyjnym – z klasy górnych 10% są odpowiedzialni za ponad 60% wszystkich publikacji powstających w tej grupie wiekowej. I są to wyniki zgodne z oczekiwaniami.

Aby oszacować nadreprezentację mężczyzn w klasach najbardziej produktywnych naukowców z perspektywy różnych dyscyplin i w ujęciu czasu stworzyliśmy Wskaźnik Względnej Obecności (WWO) dla mężczyzn. Mężczyźni są nadreprezentowani we wszystkich klasach najbardziej produktywnych naukowców, a ich nadreprezentacja rośnie wraz z przesuwaniem się w górę skali wysokiej produktywności. Wskaźnik ten maleje wraz z kolejnymi grupami wieku akademickiego: najmłodsi mężczyźni są najbardziej nadreprezentowani, a najstarsi najmniej.

Innymi słowy, we wszystkich okresach i dla trzech klas najbardziej produktywnych naukowców kobiety są niedoreprezentowane w najmniejszym stopniu w najstarszych grupach wiekowych, a w największym stopniu w najmłodszych grupach wiekowych. Potwierdza to powszechne przekonanie, że najtrudniej jest konkurować z mężczyznami młodym kobietom, a później przewaga mężczyzn stopniowo maleje (zob. analizę młodych naukowców w krajach OECD w Kwiek i Szymula 2023).

Wyniki naszych modeli ekonometrycznych prowadzą do kilku interesujących wniosków. Po pierwsze, model dla 10% najbardziej produktywnych naukowców identyfikuje trzy podstawowe predyktory, które zwiększają szanse przynależności do tej grupy: to płeć, wiek akademicki i wzorce publikacyjne. Mężczyźni mają średnio znacznie większe szanse (o 39,2%); każdy dodatkowy rok publikowania zwiększa szanse o 7,2%; a wzrost wskaźnika współpracy międzynarodowej w danym okresie o jedną jednostkę zwiększa szanse o 1%. Zatem szanse na przynależność do grupy 10% najbardziej produktywnych naukowców są średnio znacznie wyższe w przypadku mężczyzn z długą historią publikacyjną (czyli starszych) i silnie zaangażowanych w międzynarodową współpracę badawczą.

Po drugie, zatrudnienie poza grupą 10 uczelni uczestniczących dzisiaj w programie IDUB (m.in. poza UW, UJ, UAM, PW, AGH czy UMK) odgrywa tylko niewielką negatywną (ale statystycznie istotną) rolę. To wynik w pewnym sensie sprzeczny z naszymi intuicjami: spodziewaliśmy się bowiem, że najbardziej produktywni naukowcy będą pracować głównie w instytucjach największych i intensywnie prowadzących badania. A tak nie jest.

I wreszcie po trzecie, analiza efektów stałych w naszym modelu ekonometrycznym wskazuje, że z perspektywy ostatnich 30 lat coraz trudniej jest wejść do grup najbardziej produktywnych naukowców, co świadczy o rosnącej konkurencji w polskiej nauce. Najtrudniejsze jest wejście do klasy górnego 1% w ostatnim badanym okresie, czyli 2016–2021. Wszyscy publikujemy więcej, ale zarazem wyraźnie więcej publikują analizowane trzy grupy najbardziej produktywnych naukowców.

Podziękowania

Dziękujemy za wsparcie udzielone w ramach grantu MEiN NDS nr NdS/529032/2021/2021. Dziękujemy za pomoc Kristy James i Alickowi Birdowi z International Center for the Study of Research (ICSR Lab) w Dordrechcie. MK dziękuje uczestnikom seminarium, które poprowadził 27 maja 2024 w German Center for Higher Education Research and Science Studies (DZHW) w Berlinie. I wreszcie jesteśmy bardzo wdzięczni dr. Łukaszowi Szymuli za wsparcie przy pozyskiwaniu i analizie danych z bazy ICSR Lab.

Bibliografia

- Abramo G., D'Angelo C.A., & Caprasecca A. (2009). The contribution of star scientists to overall sex differences in research productivity. *Scientometrics*, 81(1), 137–156.
- Abramo G., D'Angelo C.A., Soldatenkova, A. (2017). How long do top scientists maintain their stardom? An analysis by region, gender and discipline: Evidence from Italy. *Scientometrics* 110, 867–877.

- Allison P.D. (2000). *Fixed effects regression models*. Sage.
- Fernández-Val I., M. Weidner (2016). Individual and time effects in nonlinear panel models with large N, T. *Journal of Econometrics*, 192(1), 291–312.
- Hinz J., A. Stammann, J. Wanner (2020). State dependence and unobserved heterogeneity in the extensive margin of trade. *ArXiv e-prints*.
- Kelchtermans S., Veugelers R. (2013). Top research productivity and its persistence: Gender as a double-edged sword. *Review of Economics and Statistics*, 95(1), 273–285.
- Kwiek M. (2016). The European research elite: A cross-national study of highly productive academics across 11 European systems. *Higher Education*, 71(3), 379–397.
- Kwiek M. (2018a). High research productivity in vertically undifferentiated higher education systems: Who are the top performers? *Scientometrics*, 115(1), 415–462.
- Kwiek M. (2018b). Academic top earners. Research productivity, prestige generation and salary patterns in European universities. *Science and Public Policy*. 45(1). February 2018. 1–13.
- Kwiek M. (2019). *Changing European Academics. A Comparative Study of Social Stratification, Work Patterns and Research Productivity*. London and New York: Routledge.
- Kwiek M. (2022a). *Globalna nauka, globalni naukowcy*. Warszawa: PWN.
- Kwiek M., Roszka W. (2022b). Academic vs. biological age in research on academic careers: a large-scale study with implications for scientifically developing systems. *Scientometrics*, 127, 3543–3575.
- Kwiek M., Roszka W. (2024a). Once highly productive, forever highly productive? Full professors' research productivity from a longitudinal perspective. *Higher Education*, 87, 519–549.
- Kwiek M., Roszka W. (2024b). Are Scientists Changing their Research Productivity Classes When They Move Up the Academic Ladder? *Innovative Higher Education*. Online first: <https://doi.org/10.1007/s10755-024-09735-3>
- Kwiek M., Roszka W. (2024c). Top research performance in Poland over three decades: A multi-dimensional micro-data approach. *Journal of Informetrics*, 18(4). November 2024. 101595. 1–16.
- Kwiek M., Szymula Ł. (2023). Young male and female scientists: A quantitative exploratory study of the changing demographics of the global scientific workforce. *Quantitative Science Studies*, 4(4), 902–937.
- Kwiek M., Szymula Ł. (2024). Quantifying attrition in science: a cohort-based, longitudinal study of scientists in 38 OECD countries. *Higher Education*. Online first: <https://doi.org/10.1007/s10734-024-01284-0>
- Larivière V., Gingras Y. (2010). The Impact Factor's Matthew Effect. A natural experiment in bibliometrics. *Journal of the American Society for Information Science and Technology* 61(2) 424–427.
- Stammann A. (2018). Fast and feasible estimation of generalized linear models with high dimensional k-way fixed effects. *ArXiv e-prints*.
- Turner L., Mairesse J. (2005). Individual productivity differences in public research: How important are non-individual determinants? An econometric study of French physicists' publications and citations (1986–1997). CNRS.

Tabela uzupełniająca 1. Rozkład podpróby klas najbardziej produktywnych naukowców (górných 10%, 5% i 1% pod względem produktywności) według sześcioletniego okresu, płci, typu instytucji i grupy wieku akademickiego (doświadczenia w publikowaniu)

		Górných 10%		Górných 5%		Górny 1%	
		<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
1992–1997	Kobiety	649	27,0	301	25,0	33	13,7
	Mężczyźni	1751	73,0	901	75,0	208	86,3
	Łącznie	2400	100,0	1202	100,0	241	100,0
1998–2003	Kobiety	1135	30,7	498	27,0	73	19,5
	Mężczyźni	2567	69,3	1348	73,0	301	80,5
	Łącznie	3702	100,0	1846	100,0	374	100,0
2004–2009	Kobiety	1823	33,4	808	29,6	124	22,5
	Mężczyźni	3640	66,6	1918	70,4	426	77,5
	Łącznie	5463	100,0	2726	100,0	550	100,0
2010–2015	Kobiety	2755	35,9	1207	31,4	189	24,5
	Mężczyźni	4927	64,1	2634	68,6	581	75,5
	Łącznie	7682	100,0	3841	100,0	770	100,0
2016–2021	Kobiety	3669	39,3	1653	35,4	240	25,7
	Mężczyźni	5668	60,7	3013	64,6	694	74,3
	Łącznie	9337	100,0	4666	100,0	934	100,0
1992–1997	IDUB	645	26,9	299	24,9	61	25,3
	Pozostałe	1755	73,1	903	75,1	180	74,7
	Łącznie	2400	100,0	1202	100,0	241	100,0
1998–2003	IDUB	1016	27,4	522	28,3	119	31,8
	Pozostałe	2686	72,6	1324	71,7	255	68,2
	Łącznie	3702	100,0	1846	100,0	374	100,0
2004–2009	IDUB	1521	27,8	756	27,7	163	29,6
	Pozostałe	3942	72,2	1970	72,3	387	70,4
	Łącznie	5463	100,0	2726	100,0	550	100,0
2010–2015	IDUB	2189	28,5	1089	28,4	228	29,6
	Pozostałe	5493	71,5	2752	71,6	542	70,4
	Łącznie	7682	100,0	3841	100,0	770	100,0
2016–2021	IDUB	2564	27,5	1281	27,5	271	29,0
	Pozostałe	6773	72,5	3385	72,5	663	71,0
	Łącznie	9337	100,0	4666	100,0	934	100,0
1992–1997	0–9 lat	763	31,8	317	26,4	44	18,3
	10–19 lat	860	35,8	439	36,5	70	29,0
	20–29 lat	642	26,8	366	30,4	101	41,9

	30 lat i więcej	135	5,6	80	6,7	26	10,8
	Łącznie	2400	100,0	1202	100,0	241	100,0
1998-2003	0-9 lat	1103	29,8	390	21,1	52	13,9
	10-19 lat	1224	33,1	624	33,8	108	28,9
	20-29 lat	999	27,0	595	32,2	145	38,8
	30 lat i więcej	376	10,2	237	12,8	69	18,4
	Łącznie	3702	100,0	1846	100,0	374	100,0
2004-2009	0-9 lat	1404	25,7	546	20,0	56	10,2
	10-19 lat	2014	36,9	1008	37,0	178	32,4
	20-29 lat	1167	21,4	652	23,9	158	28,7
	30 lat i więcej	878	16,1	520	19,1	158	28,7
	Łącznie	5463	100,0	2726	100,0	550	100,0
2010-2015	0-9 lat	1708	22,2	649	16,9	73	9,5
	10-19 lat	3239	42,2	1592	41,4	285	37,0
	20-29 lat	1478	19,2	857	22,3	200	26,0
	30 lat i więcej	1257	16,4	743	19,3	212	27,5
	Łącznie	7682	100,0	3841	100,0	770	100,0
2016-2021	0-9 lat	1893	20,3	751	16,1	84	9,0
	10-19 lat	3862	41,4	1948	41,7	358	38,3
	20-29 lat	2298	24,6	1226	26,3	302	32,3
	30 lat i więcej	1284	13,8	741	15,9	190	20,3
	Łącznie	9337	100,0	4666	100,0	934	100,0

Tabela uzupełniająca 2. Efekty stałe w modelach

Okres, Dyscyplina/Klasa	Górnych 10%	Górnych 5%	Górny 1%
1992-1997	-3,728	-4,835	-7,218
1998-2003	-3,827	-4,949	-7,343
2004-2009	-3,896	-5,026	-7,445
2010-2015	-3,940	-5,075	-7,517
2016-2021	-4,092	-5,241	-7,717
AGRI	0,186	0,212	0,273
BIO	-0,072	-0,075	-0,087
CHEM	-0,196	-0,223	-0,286
CHEMENG	0,125	0,153	0,260
COMP	0,069	0,087	0,090
EARTH	-0,098	-0,104	-0,123
ENER	0,386	0,434	0,598
ENG	0,113	0,119	0,142

ENVIR	0,359	0,407	0,525
MATER	0,048	0,048	0,034
MATH	-0,201	-0,235	-0,365
MED	0,145	0,166	0,206
NEURO	0,041	0,068	0,130
PHARM	0,072	0,088	0,160
PHYS	-0,669	-0,770	-0,942

Reguła 10/50. Analiza dorobku 150 tysięcy polskich naukowców pokazuje trwałość systemowych nierówności w nauce

Przeprowadziliśmy analizę udziału najbardziej produktywnych polskich naukowców w krajowej produkcji publikacyjnej z perspektywy czasowej (1992–2021) i dyscyplinarnej (15 dyscyplin STEMM). Górny 1% naukowców odpowiadał średnio za 10% krajowej produkcji publikacyjnej (reguła 1/10), a górnych 10% naukowców odpowiadało za prawie 50% tej produkcji (reguła 10/50). Wyniki naszych modeli ekonometrycznych pokazują, że szanse na przynależność do grupy najbardziej produktywnych naukowców są średnio znacznie wyższe w przypadku mężczyzn, naukowców z długą historią publikacyjną i szeroko zaangażowanych w międzynarodową współpracę badawczą. Mężczyźni są nadreprezentowani we wszystkich trzech badanych klasach najbardziej produktywnych naukowców (1%, 5% i 10%). Zbiór danych zawierał $N=152\,043$ naukowców i ich $N_{\text{art}} = 587\,558$ artykułów z okresu 1992–2021 indeksowanych w bazie Scopus.

Słowa kluczowe: kariera naukowa, najbardziej produktywni naukowcy, badanie podłużne (longitudinalne), prestiż czasopisma, produktywność publikacyjna, logitowe uogólnione modele liniowe, efekty stałe, Polska

The 10/50 rule. An analysis of research output of 150,000 Polish scientists shows permanent systemic inequalities in science

We conducted an analysis of the contribution of the most productive Polish scientists to national publication output from a temporal (1992–2021) and disciplinary (15 STEMM disciplines) perspectives. The top 1% of scientists accounted on average for 10% of national publication output (the 1/10 rule), and the top 10% of scientists accounted for almost 50% of this output (the 10/50 rule). The results of our econometric models show that the odds of membership in the top performers classes are significantly higher for men, scientists with a long publication history, and extensively involved in international research collaboration. Men are overrepresented in all three classes of most productive scientists (1%, 5%, and 10%) examined. The dataset included $N=152\,043$ Polish scientists and their $N_{\text{art}} = 587\,558$ research articles from 1992–2021 indexed in Scopus.

Key words: scientific career, highly productive scientists, longitudinal study, journal prestige, productivity, logit generalized linear models, fixed effects, Poland

