Surprising Patterns of Changing Productivity Classes. A Longitudinal Study of 320,000 Scientists

University of Oxford, CGHE Center for Global Higher Education, May 14, 2026

INSTITUTE FOR ADVANCED STUDIES IN SOCIAL SCIENCES AND HUMANITIES

Professor Marek Kwiek

Institute for Advanced Studies in Social Sciences and Humanities (IAS), Director

UNESCO Chair in Institutional Research and Higher Education Policy

University of Poznan, Poland

kwiekm@amu.edu.pl

Twitter: @Marek_Kwiek

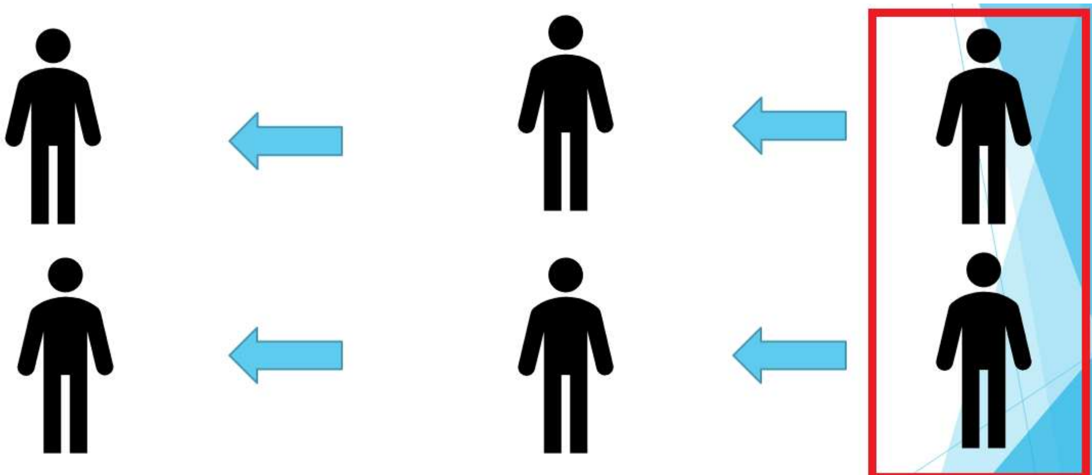
2. Overview

- **Focus on persistence** in research productivity over the course of an individual's **entire** scientific career.
- Mobility (and immobility) between productivity classes, lifetime.
- We track "**late-career**" scientists—scientists with **at least 25 years of publishing experience** (N=320,564)—in 16 STEMM and social science disciplines from 38 OECD countries, for up to 50 years
- Our sample is 79% of all late-career scientists, globally.
- We examine the details of their **productivity mobility patterns as early-career (5-14 years), mid-career (15-24 years), and late-career (at least 25 years) scientists.**
- **Methodologically**: we turn a large-scale, curated, structured **publication and citation bibliometric dataset** (Scopus raw data) into

- a global, comprehensive, multidimensional, and **longitudinal data source** for research on careers in science

3. Simple Longitudinal Study Design.

We study how scientific careers **change over time** (Menard, 2002; Rowland, 2014; Ruspini, 1999).
The same individuals are tracked **over the decades** of their publishing careers.
Microdata about each individual scientist are used.
Retrospective examination is applied (looking back).
Three stages used: Late-career, mid-career, and early career (not: beginners)



4. Collaborative Reseach: My Team

- Research with Dr. Lukasz Szymula from Poznan CPPS Team.
- In Aaron Clauset's Lab (2023), and Cassidy Sugimoto's Lab (2025).

5. Three (Competitive) Small-Scale Studies

- There are **three small-scale <u>longitudinal</u> single-nation studies similar to ours. A few hundred to 3000 cases only.**
- First, **for 497 French physicists, Turner and Mairesse (2005)** showed that 66% of the most productive researchers („quartile 1 scientists”) remained as such for the period 1986–1997.
- Second, in **a study of a single Belgian university, Kelchtermans and Veugelers (2013)** discussed 1,040 scientists from 1992–2001; they studied how researchers switch between productivity categories over time, showing strong support for a accumulative processes.
- Finally, **Abramo et al. (2017) studied Italian scientists in three consecutive four-year periods of 2001-2012**; they identified 2,883 top performers in the first period. One-third of top performers retained their top ranking for three consecutive periods, half retained their „stardom” for two periods (35% and 55%, respectively).


Methodology

7. Individuals, Not Publications. Microdata, not Aggregated Data

- In this research, we move **from individual publications (and their properties)** to **individual scientists (and their characteristics)** as a unit of analysis.
- We use curated, commercial BiG Data – raw Scopus dataset.
- We construct **individual lifetime publication and citation histories** for every late-career scientist in our sample. We restrict our research to 16 STEMM and social science disciplines.
- In our context, "**late-career**" scientists are defined as scientists with **at least 25 years** of publishing experience (recorded in Scopus database)
- We **explore mobility between the 10 individual productivity classes** (constructed according to **the 10 decile-based classes**) throughout long academic careers.

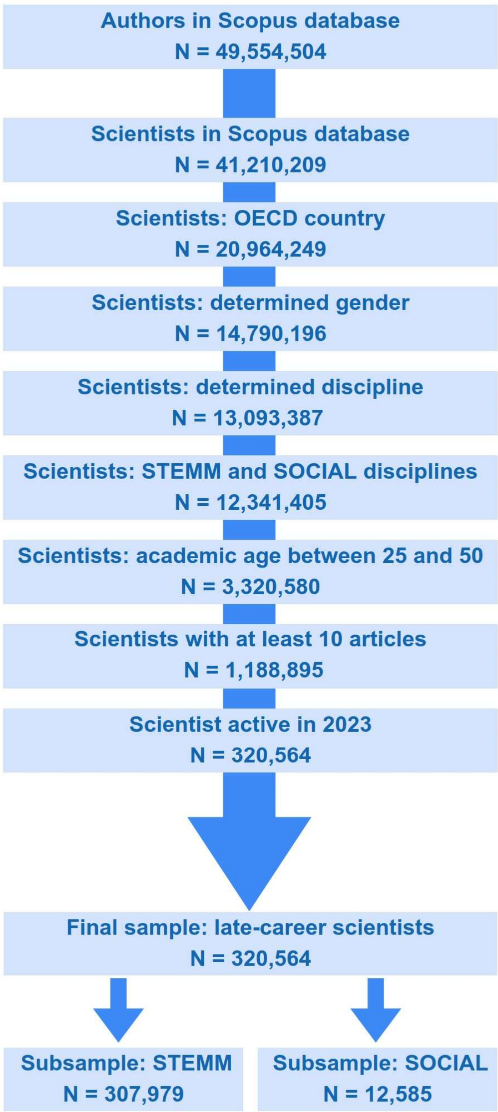
8. Dataset

Our final sample included all late-career scientists who:
- were research active in 2023
- had at least 25 years of publishing experience
- located in 16 STEMM and social science disciplines and
- come from 38 OECD countries
- N=320,564 scientists with N=16,345,891 research articles;

For our calculations, we utilized the Scopus raw database (ICSR Lab) dated March 29, 2024.

We defined:

Non-occasional scientists (at least ten journal articles);
Country affiliation as an OECD country (modal value);
Gender (binary: male or female, probability at least 0.85);
Discipline as STEMM (all cited references lifetime; modal value) and SOCIAL. Defined via Lifetime References.



9. Productivity Classes, Current & Retrospective Distribution

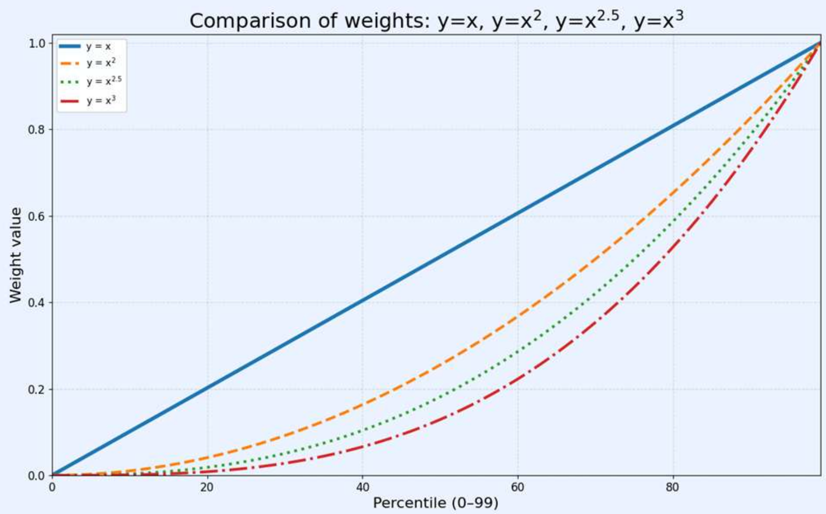
- Academic lives of our late-career scientists were **retrospectively divided** into three stages:
  - early-career stage; mid-career stage, and late-career stage.
- We analyzed their **current five-year publishing behavior (2019-2023) and looked back** into their past publishing behavior - to examine how they may have changed their productivity classes.
- **At each career stage, current late-career scientists showed their annual individual publishing productivity.**
- **Productivity was calculated for the recent five-year period and for two earlier periods: when they were early-career scientists and mid-career scientists.**
- Our analyses are based on the idea of subsequent distributions of scientists into **classes**: we Focus on distributions of scientists, not publication numbers.

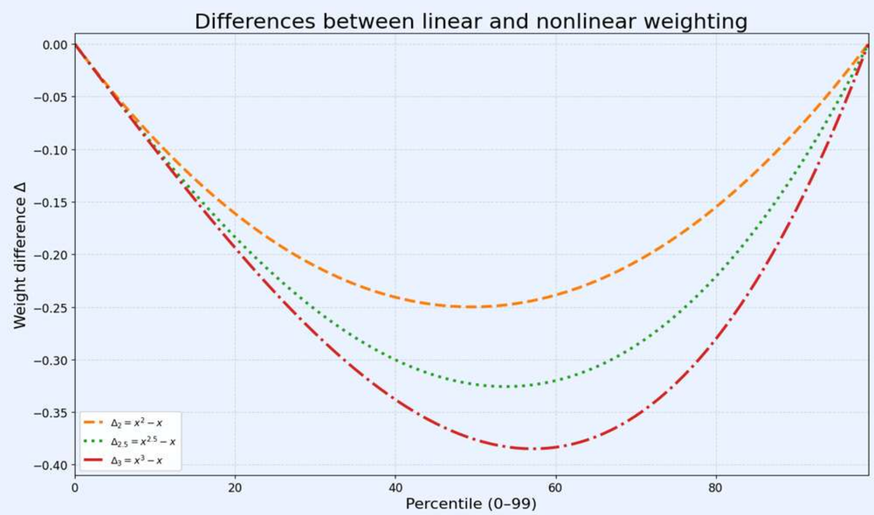
10. JPNP: Journal Prestige Normalized Productivity (1)

- We used a **journal prestige-normalized, full counting method** of calculating productivity.
- Our JPNP refers to both **quantity and quality** of globally indexed publications at the level of individuals (as opposed to quantity only in journal **non-normalized** approaches).
- Prestige normalization refers to **journal percentile ranks used in Scopus database (CiteScore ranking, range: 1-99**).
- Top journals – usually 90-99th journal percentile ranks.
- JPNP highlights the **difference in average scholarly efforts** between preparing and revising publications in generally less selective - and more selective journals (different **peer review procedures, peer reviewers and acceptance rates**).
- **In JPNP the weight of publications depends on their location in a vertically stratified system of academic journals** (journal stratification, see Hammarfelt, 2017; Heckman & Moktan, 2018; Kwiek 2021; Lindahl, 2018).

11. JPNP: Journal Prestige-Normalized Productivity (2)

- **JPNP is nonlinear indicator** of individual publication productivity: publications in top-tier journals have much higher weigth than publications in low-tier journals.
- Traditional linear and non-prestige normalized productivity measures fail **to capture the contemporary hierarchical structure of journal prestige (in the global publishing system).**
- JPNP yields a **realistic distribution of productivity** and **reduces the noise** generated in productivity computations by low-prestige publications.
- The indicator relies on **journal prestige (1-99th percentiles) as a (proxy) signal of quality.**
- This is an ***ex ante* measure rather than an *ex post* measure** (such as field-normalized **citations, before/after**): it captures **expected visibility at the journal level.**
- JPNP **shifts the interpretation of productivity away from publication activity alone (pure volume, counts**) toward productivity embedded in **the global journal prestige hierarchy**.



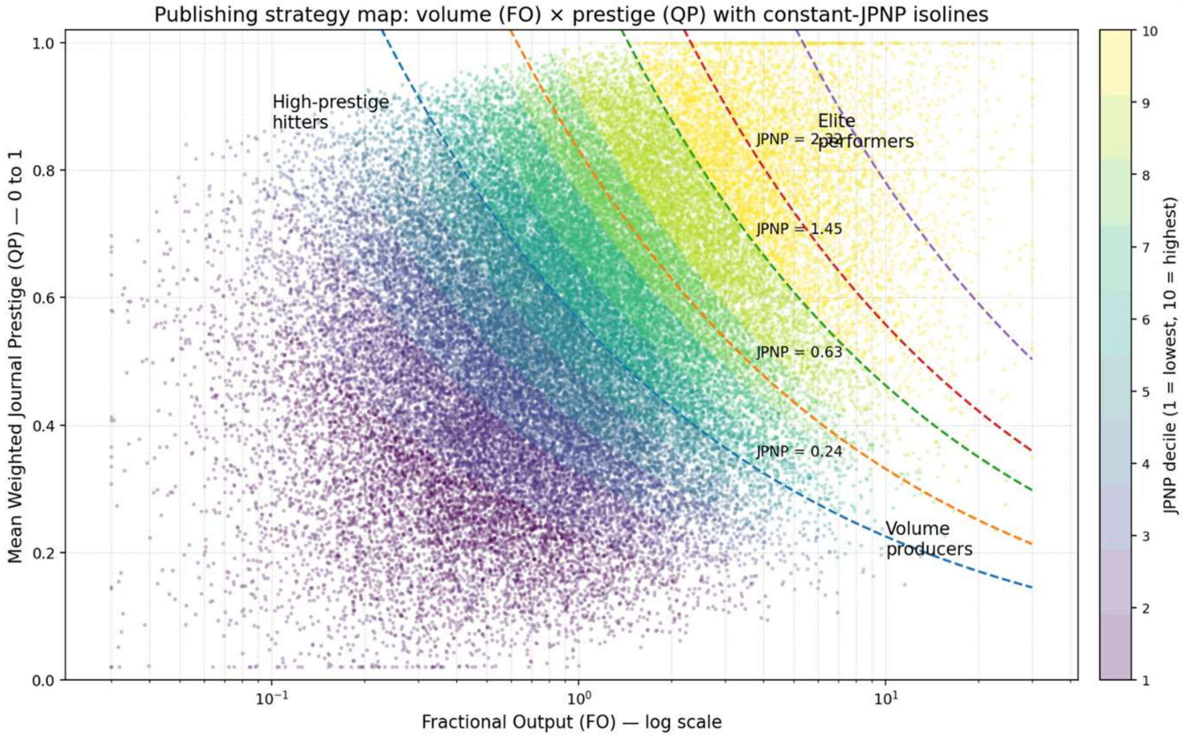


- **12. Table. Prestige weight values for selected percentiles in four weighting schemes:** linear ($y = x$) and nonlinear ($y = x^2$, $y = x^{2.5}$, $y = x^3$). Column *x (normalized)* shows the journal rank percentile (0–99) scaled to the range [0,1].
- The table shows that **as the exponent (*k*) increases, the differences between journals become more pronounced: low and medium percentiles are increasingly depreciated, while the upper tail (top 10%, top 5%, top 1%) gains relative importance.**

Percentile	x (normalized)	$y = x$	$y = x^2$	$y = x^{2.5}$	$y = x^3$
0	0.0000	0.0000	0.0000	0.0000	0.0000
10	0.1010	0.1010	0.0102	0.0032	0.0010
20 (bottom 20%)	0.2020	0.2020	0.0408	0.0183	0.0082
30	0.3030	0.3030	0.0918	0.0505	0.0278
40	0.4040	0.4040	0.1632	0.1038	0.0660
50 (median)	0.5051	0.5051	0.2551	0.1813	0.1288
60	0.6061	0.6061	0.3673	0.2860	0.2226
70	0.7071	0.7071	0.4999	0.4204	0.3535
80	0.8081	0.8081	0.6530	0.5870	0.5277
90 (top 10%)	0.9091	0.9091	0.8264	0.7880	0.7513
95 (top 5%)	0.9596	0.9596	0.9208	0.9020	0.8836
99 (top 1%)	1.0000	1.0000	1.0000	1.0000	1.0000

13. JPNP - Journal Prestige-Normalized Productivity (4)

- **The nonlinear transformation $x^{2.5}$ can be read as a description of the structure of contemporary published science - rather than merely a mathematical device.**
- The global publishing system is hierarchical. And marginal prestige differences at the top translate into disproportionate reward differences (Kwiek 2018; 2024).
- Linear prestige models cannot capture this stratification.
- **JPNP does not create inequalities; it reveals them**, because inequality is embedded in the structure of journal prestige.
- A natural development direction for JPNP is **hybrid** solutions combining *ex ante* prestige weighting with **field-normalized *ex post* citation measures** (which we are currently developing under the label of JPNP 2.0).
- **JPNP takes weight away from publications at the bottom and middle of the Journal prestige distribution.**
- **The dominance of the upper tail of journals is increased.**



Note: JPNP is illustrated as FO × QP^k (k = 2.5) to visualize constant-score trade-offs.
In empirical applications, JPNP is computed at the publication level and then aggregated.

- **14. The same level of JPNP can be achieved through different publication <u>strategies</u>**: increasing output volume, raising journal prestige or combining both.
- Isolines depict **the trade-off between quantity and prestige at constant JPNP**, and decile shading reveals increasing concentration in the region of high output and high journal prestige (a rare publication elite).
- The figure shows distinct locations for **high-prestige hitters** (low volume, high prestige), **volume producers** (high volume, moderate prestige), and **elite performers** (high volume and high prestige).

15. Methodological Approach – Individual Attributes

- **Gender determination**
  Based on first name, last name, and dominant country in Year 1 (Namsor, prob≥0.85). Binary. Manual curation (500) using website photos.
- **Discipline determination**
  Each cited reference (N=1.43 billion) is assigned the journal's discipline. For each author, references are counted by discipline (excluding "multidisciplinary"). The discipline with the highest references (modal value) is selected. Lifetime. In Polish Studies of 30 yrs – disciplines for 5 periods.
- **Determining the country of affiliation**
  For each author, we count countries listed across publications. The country with the highest frequency (modal value) is selected.
- **Determining scientists' nonoccasional status**
  A nonoccasional scientist has at least ten published articles.
- **Determining academic age / publishing experience**
  For each author, the **year of the first** publication (of any type) was determined.

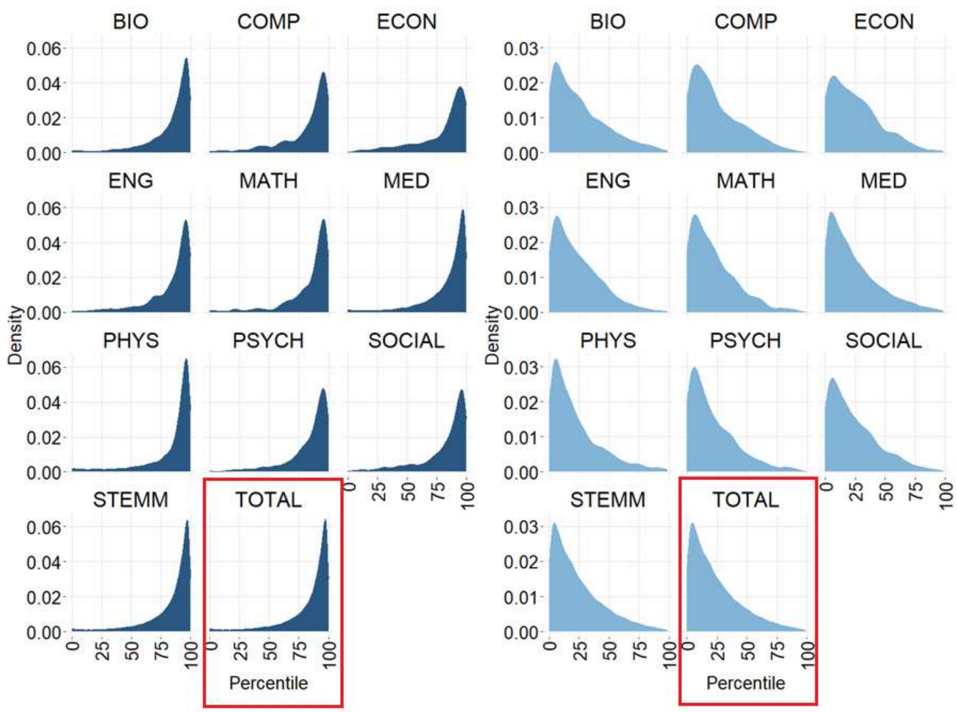
List of STEMM disciplines

16 STEMM disciplines (All Science Journal Classification, ASJC): **AGRI**, agricultural and biological sciences; **BIO**, biochemistry, genetics, and molecular biology; **CHEMENG**, chemical engineering; **CHEM**, chemistry; **COMP**, computer science; **EARTH**, earth and planetary sciences; **ENER**, energy; **ENG**, engineering; **ENVIR**, environmental science; **IMMU**, immunology and microbiology; **MATER**, materials science; **MATH**, mathematics; **MED**, medicine, **NEURO**, neuroscience; **PHARM**, pharmacology, toxicology, and pharmaceutics; and **PHYS**, physics and astronomy.

Social: **BUS** Business, **ECON**, Economics and Econometrics, and **PSYCH** Psychology

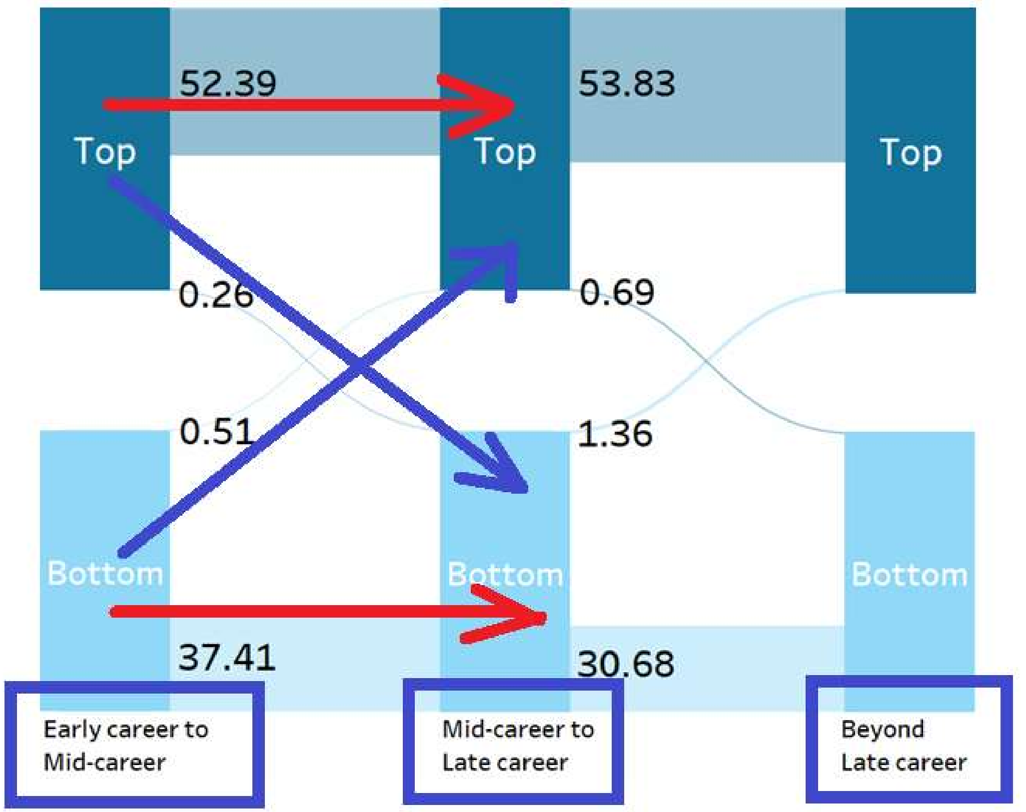

INVESTIGATION
RESULTS

17. Kernel density plots, initial percentile distribution of top performers and bottom performers, by selected discipline.
Late-career scientists.
Where do top performers come from?
Top performers (decile 10) come predominantly from deciles 8-10 at earlier stage
Bottom performers (decile 1) come predominantly from deciles 1-3 at earlier stage

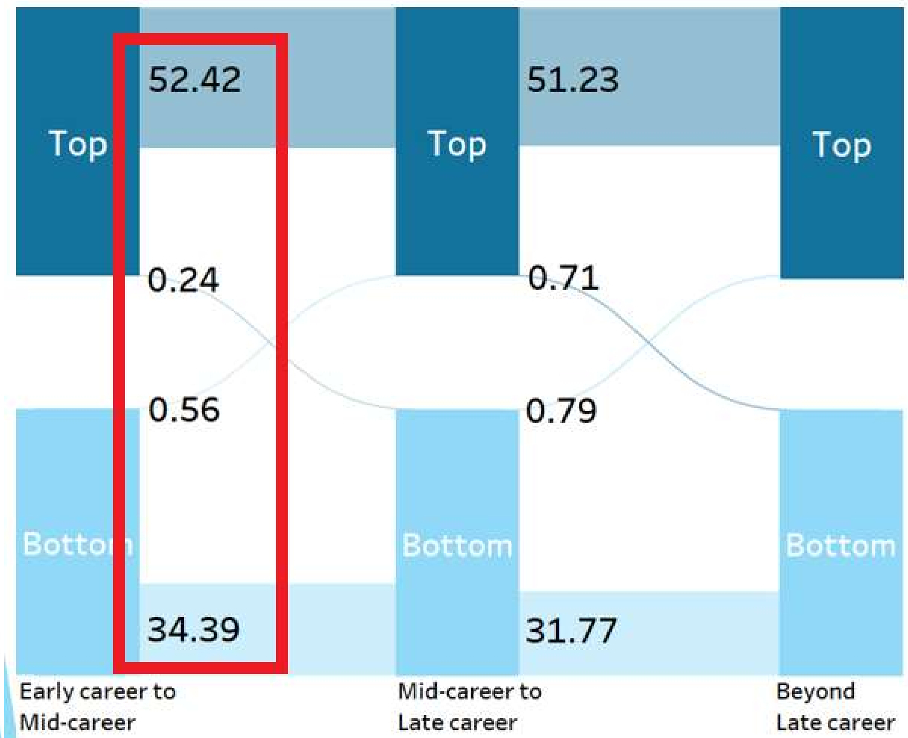


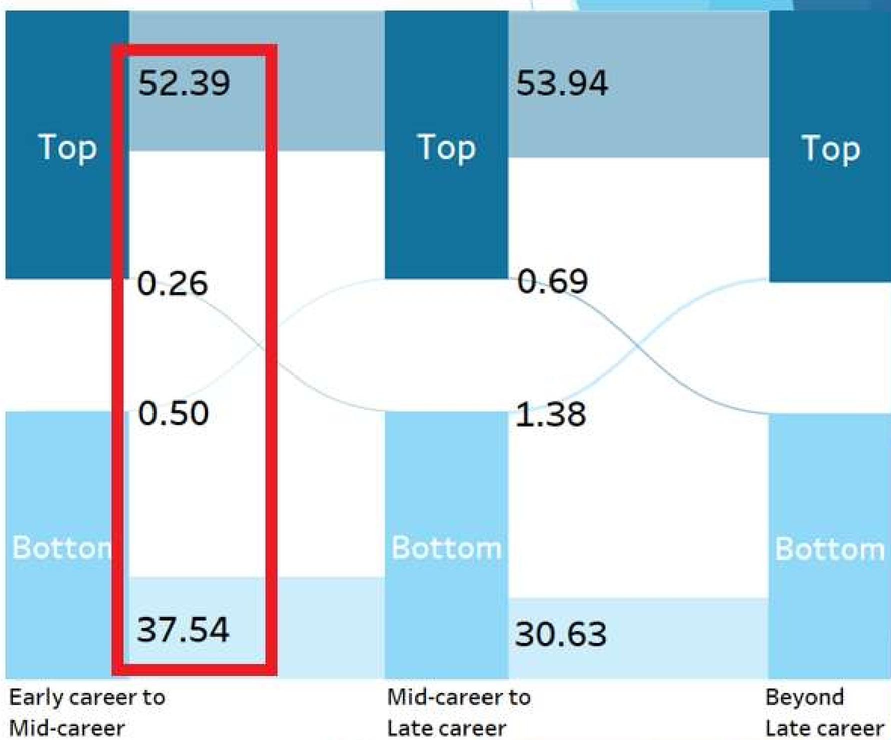
18. Sankey Diagrams: Understanding and Quantifying Mobility / Immobility (flows) between Productivity Classes

- The Sankey diagrams are visual guides to better understand **the concept of scientists' mobility across productivity** classes throughout their **careers (TOTAL).**
- This diagram illustrates the movement of scientists **between productivity deciles at different career stages**: early career (left: top and bottom), mid-career (middle: top and bottom), and late career (right: top and bottom).
- Our focus is on **horizontal top-to-top** mobility as well as the transitions involving **extreme upward mobility** from bottom-to-top.
- Other productivity deciles removed for clarity (N = 320,564) (percentages, top class, and bottom class 100% each)



19. All social science disciplines combined (SOCIAL) (N = 12,585)
All STEMM academic disciplines combined (N = 307,979)



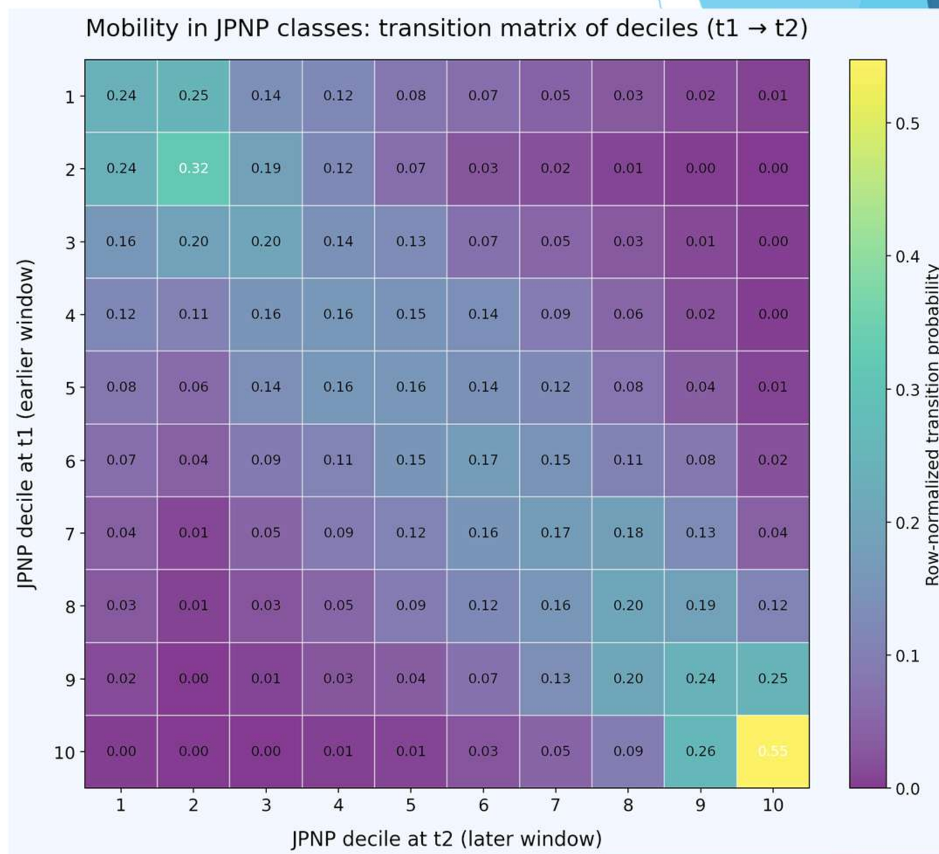


20. Mobility between top- (decile 10) and bottom- (decile 1) productivity classes while moving up from the early-career stage to mid-career stage and from the mid-career to late-career stage by all academic disciplines combined (N = 320,564, top panel), social science disciplines combined (N = 12,585, middle panel), and STEMM disciplines combined (N = 307,979, bottom panel) (frequencies and percentages)

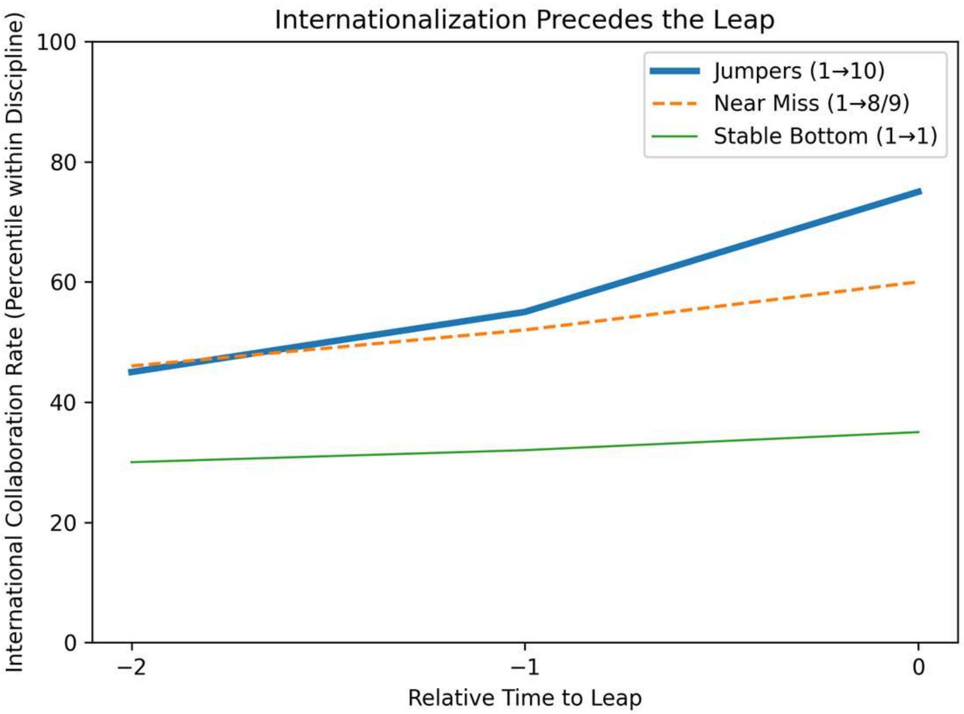
Career stage (transition from)	Initial productivity decile	Career stage (transition to)	Target productivity decile	Number of scientists in transition	Number of scientists in productivity class	%
TOTAL (ALL DISCIPLINES COMBINED)						
Early career	Bottom	Mid-career	Bottom	11,996	32,063	37.41
Early career	Bottom	Mid-career	Top	162	32,063	0.51
Early career	Top	Mid-career	Bottom	82	32,063	0.26
Early career	Top	Mid-career	Top	16,799	32,063	52.39
Mid-career	Bottom	Late career	Bottom	9,836	32,063	30.68
Mid-career	Bottom	Late career	Top	436	32,063	1.36
Mid-career	Top	Late career	Bottom	222	32,063	0.69
Mid-career	Top	Late career	Top	17,261	32,063	53.83
ALL SOCIAL SCIENCE DISCIPLINES						
Early career	Bottom	Mid-career	Bottom	433	1,259	34.39
Early career	Bottom	Mid-career	Top	7	1,259	0.56
Early career	Top	Mid-career	Bottom	3	1,259	0.24
Early career	Top	Mid-career	Top	660	1,259	52.42
Mid-career	Bottom	Late career	Bottom	400	1,259	31.77
Mid-career	Bottom	Late career	Top	10	1,259	0.79
Mid-career	Top	Late career	Bottom	9	1,259	0.71
Mid-career	Top	Late career	Top	645	1,259	51.23
ALL STEMM DISCIPLINES						
Early career	Bottom	Mid-career	Bottom	11,563	30,804	37.54
Early career	Bottom	Mid-career	Top	155	30,804	0.50
Early career	Top	Mid-career	Bottom	79	30,804	0.26
Early career	Top	Mid-career	Top	16,139	30,804	52.39
Mid-career	Bottom	Late career	Bottom	9,436	30,804	30.63
Mid-career	Bottom	Late career	Top	426	30,804	1.38
Mid-career	Top	Late career	Bottom	213	30,804	0.69
Mid-career	Top	Late career	Top	16,616	30,804	53.94

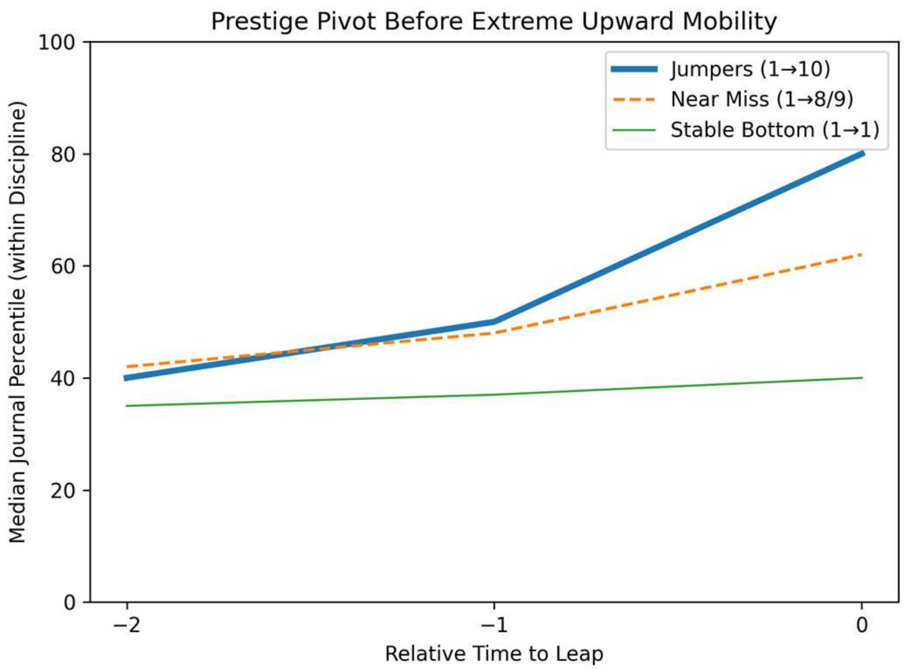
21. Jumpers-Up: Only One Economist, One Immunologist… in 38 OECD Countries!

- Extreme upward mobility (1→10 decile transitions) is empirically rare in global science, **affecting fewer than 1% of late-career scientists** in our dataset of 320,564 individuals.
- Only a small fraction of these scientists moved up from the lowest three deciles, with just **162 making a significant leap from decile 1 to decile 10** (0.51%) and 232 from decile 2 to decile 10 (0.72%).
- **We have full lifetime biographical and publishing profiles of every scientist, including these few hundreds of outliers**.
- Jumpers-Up across disciplines: from 0.26% in ECON to 1.33% in PHYS.
- **Only one economist** (0.26%) and **one immunologist** (0.32%) made the leap from decile 1 to decile 10 (out of 385 and 315)
- Only one in 38 OECD countries!

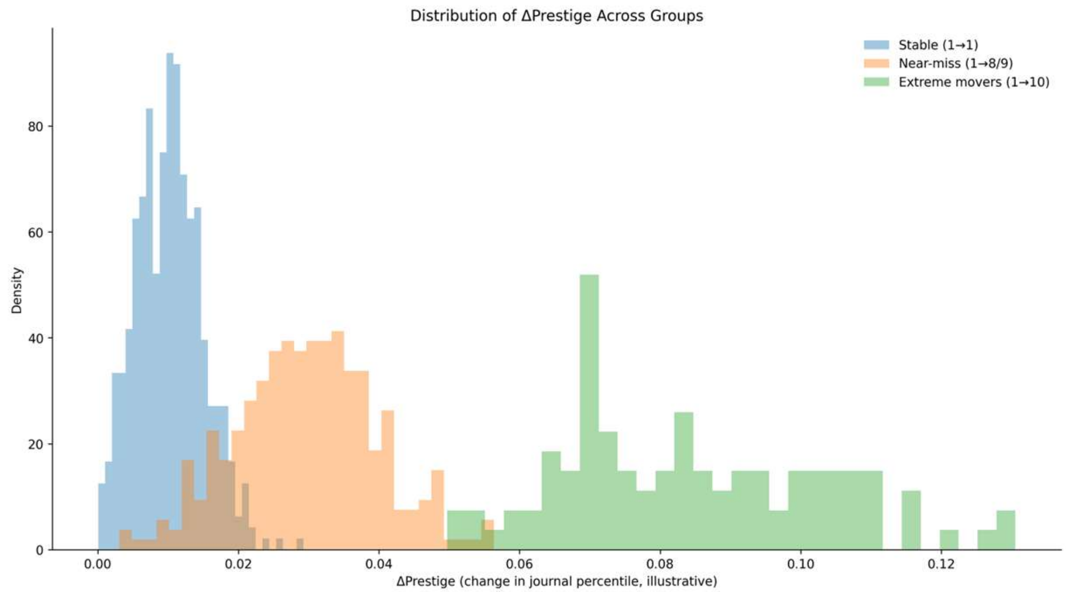


22. Why Some Scientists Become Jumpers-Up?





- **Mechanisms, not descriptions** (**Jumpers-Up vs. Non-Jumpers**)
- **Matching**: cases and controls (**near-misses 1 to 8/9, stable bottoms 1 to 1, typical jumpers 1 to 10**). Here: Internationalization in teams (Left) and Jump in Journal prestige (Right).
- While our prior analyses demonstrated **strong path dependence** in productivity classes, we ask now a different question: **what structural changes precede the rarely observed upward leaps?**



23. **Various options… what may be happening to Jumpers vs. Near-Misses and Stables?**

Change in journal prestige for various segments of scientsts (Right)

Prestige jump: moving into much higher-ranked journal segments
Internationalization spike: sharp rise in international co-authorship
Threshold effect: mobility happens only after crossing a critical prestige/network level
Pre-jump burst: short, concentrated acceleration right before the move
Network repositioning: new elite collaborators/affiliations that re-embed the scientist in the global network

24. Logistic regression statistics: odds ratio estimates of membership in the class of global top productive mid-career scientists (the top 10%, separate regressions for each academic discipline) (N = 320,564)

Variables	AGRI	BIO	BUS	CHEM	COMP	EARTH	ECON	ENG
R^2	0.25	0.22	0.21	0.26	0.18	0.26	0.23	0.20
Male	1.32***	1.86***	1.61***	1.30***	1.07***	1.29***	1.03*	0.78
Avg. FWCI 4y Early	1.53***	1.14***	1.04***	1.37***	1.05***	1.24***	1.12***	1.16***
Inter. Collab. Rate Early	1.01***	1.01***	1.01***	1.01***	1.00***	1.01***	1.01***	1.01***
Median Team Size Early	1.06***	1.01***	1.31***	0.99***	1.12***	1.05***	1.35***	0.96**
Top Early	15.39***	15.87***	15.10***	17.81***	14.64***	17.17***	17.49***	14.33
Constant	0.02***	0.02***	0.02*	0.02*	0.03*	0.02**	0.02*	0.06

	ENVIR	IMMU	MATER	MATH	MED	NEURO	PHYS	PSYCH
R^2	0.26	0.21	0.24	0.28	0.22	0.25	0.34	0.25
Male	1.21	2.12***	1.41*	1.05*	1.30***	2.27***	0.95	1.36**
Avg. FWCI 4y Early	1.27***	1.22***	1.41***	1.23***	1.02***	1.37***	1.05***	1.08***
Inter. Collab. Rate Early	1.01***	1.01***	1.01***	1.00***	1.01***	1.00***	1.02***	1.01***
Median Team Size Early	1.01***	1.02***	1.06***	1.40***	1.08***	1.16***	1.17***	1.25***
Top Early	18.16	11.68***	13.53**	22.97***	16.39***	14.25***	11.92**	18.52**
Constant	0.02	0.02**	0.02	0.02	0.03***	0.01**	0.01	0.02

* $p<0.05$, ** $p<0.01$, *** $p<0.001$


CONCLUSION

26. The Early Global Productivity Distribution Persists Over Time. For Decades

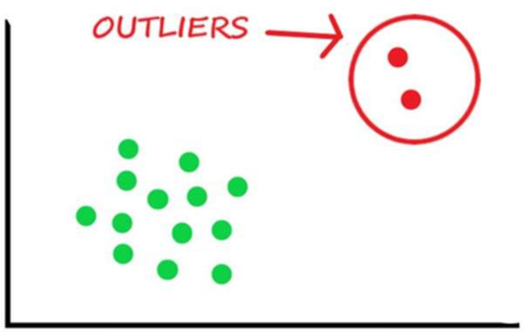
- We suggest that, **relatively early on in scientific careers**, the productivity distribution within the global science profession at its two extremes (top 10% and bottom 10%) is already largely settled.
- The early global distribution persists over time: for years and decades.
- Exceptions are very rare:
  - **global bottom performers** almost never become **global top performers** (our Jumpers-Up), and
  - **global top performers** almost never become **global bottom performers** (our Droppers-Down).

27. Prior Productivity Determines Future Productivity - Credibility Cycle in Careers

- Why do **prior productivity class memberships** (top, bottom), to a large extent, determine later class memberships (top, bottom)?
- There are two explanations, we can speculate.
- 1. Previous research has shown that **the distribution of productivity among scientists has always been highly skewed** (Abramo et al., 2017; Albarrán et al., 2011; David, 1994). A minority of scientists have always been responsible for the majority of publications (Allison, 1980; Ruiz-Castillo & Costas, 2014; Xie, 2014; Kwiek 2016).
- 2. **Higher (prestigious) publishing productivity generally leads to new research funding**, as **the credibility cycle in academic careers** shows (Latour & Woolgar, 1986).
- In this cycle, **research published in prestigious journals (quantity, quality) is converted into recognition**; successful grant applications are converted into new equipment, arguments, and articles.

28. Probabilities, Patterns & Outliers

- Our research has re-confirmed **the power of very strong track record** as opposed to **very weak track record in science** (whenever individual scientists are assessed by research funding panels, academic promotion committees etc.).
- A retrospective look at individual careers – useful in assessing the future potential of individuals.
- For a variety of reasons, the **probability** of **past global top performers becoming global top performers in the future is very high**. Their probability of becoming global bottom performers is marginal.
- At the same time, **the chances of global bottom performers** to reach the productivity levels achieved by their top-performing colleagues in the very same career stages and within their disciplines (Jumpers-Up) are **marginal**.
- **Catching up with the global top performers just does not happen**, except for a few **outliers**. (And in some system, as in Poland, it does not happen at all: The chances we have computed for Poland are 0%; Kwiek & Roszka, 2024b).

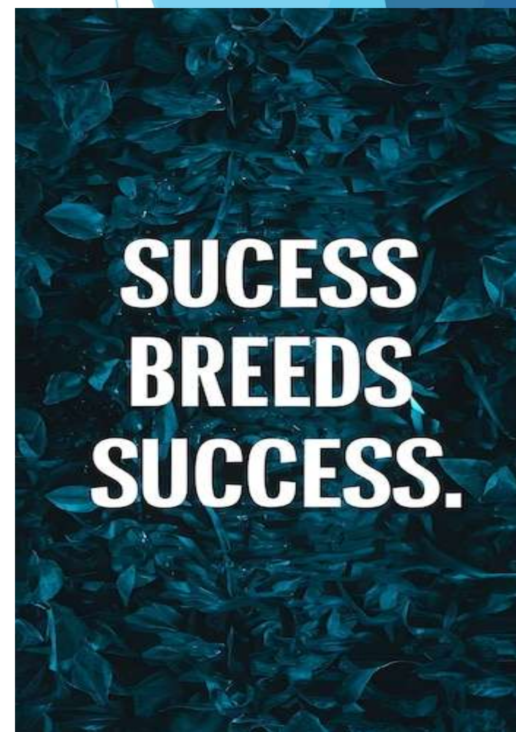


29. Implications: institutions consist of individuals. Hiring and promotions matter!

- We show that **scientists are heavily locked-in early on** in their careers in productivity classes.
- Therefore hiring and promotion decisions made at the level of **departments** have **long-lasting effects on productivity of institutions** for many years.
- **The chances of institutions that hire and promote predominantly highly productive** scientists to have **highly productive faculty are very high.**
- **Hiring and promoting bottom productive scientists** in fact means **tacit institutional agreement** to have bottom productive scientists for many years.
- **Some scientists** in both STEMM and SOCIAL clusters of disciplines **tend to be highly productive for years and decades**; and their **identification** contributes heavily to **overall institutional achievements in research**.

30. Traditional Theories of Research Productivity Revisited, Large-Scale and Longitudinal Data

- **Persistent productivity stratification:** emerges from our individual micro-level analyses as a powerful feature of global science.
- Our data and analyses confirm **what traditional productivity theories have been claiming for decades**, albeit by using **small-scale interviews and surveys** (Cole & Cole, 1973; Hermanowicz, 2012; Leišytė & Dee, 2012; Merton, 1973).
- Our Big-N study supports past small-N Studies!
- Our study shows that **success breeds success** (as in "**cumulative advantage" theory of research productivity**). Some scientists **will always be globally highly performing** while others will always be **globally low performing** (as in the "**sacred spark" theory of research productivity**).
- The power of structured Big Data shown again, following our 2024 HE study on **attrition in science** (50% within a decade, 75% within two decades).
- More in-depth research needed: **large-scale surveys** and **interviews** needed!




THANK YOU!