Megaprojects, Digital Platforms, & Productivity: Evidence from the Human Brain Project

Ann-Christin Kreyer[1] and Lucy Xiaolu Wang[2,1,3]

[1] Max Planck Institute for Innovation and Competition
[2] University of Massachusetts Amherst
[3] Canadian Centre for Health Economics

ASHEcon 2024, San Diego, 2024.6.17

Productivity slowdown, institutions, and rising AI

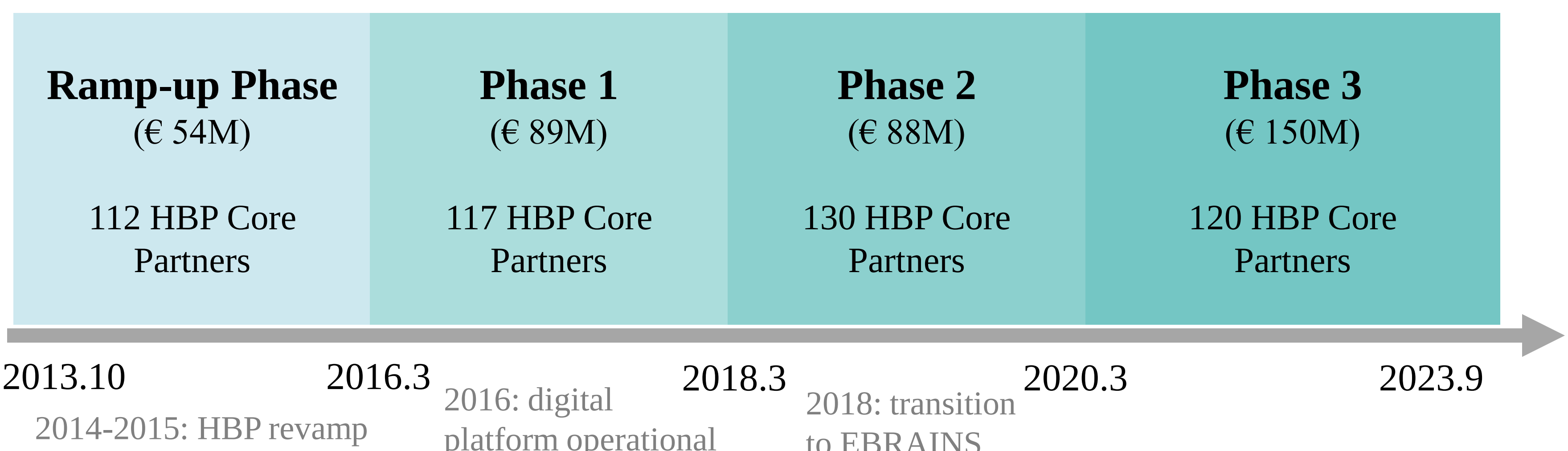
- Declining productivity in (biomedical) science & increasing challenge to reach the knowledge frontier (Jones 2009; Bloom et al. 2020)
- Important to create non-market incentives to accelerate long-term, large-scale R&D in multi-disciplinary research
- Particularly relevant & urgent for advancing neurosciences
  - Leading causes of death/disability, expensive treatment, lack of cure
- AI showed potential in life science & gained rising interests. But integrating AI with lab science is not always natural/easy
- This paper: how does a ten-year megaproject affect AI-brain sciences?

Human Brain Project

European Commission

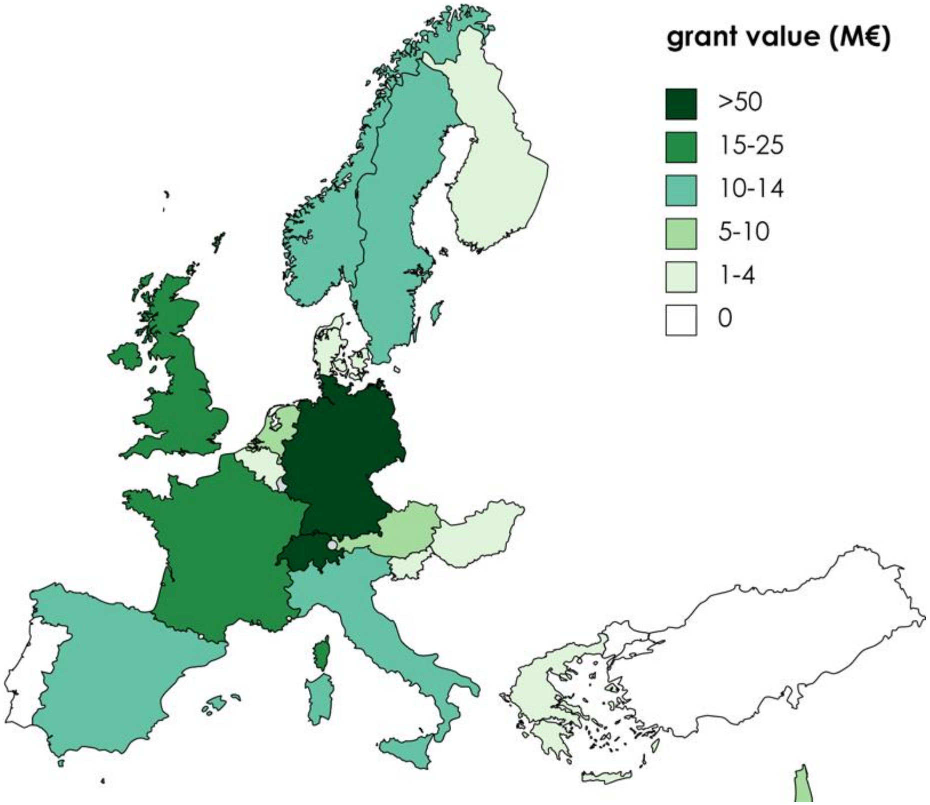
This paper: the Human Brain Project (HBP)

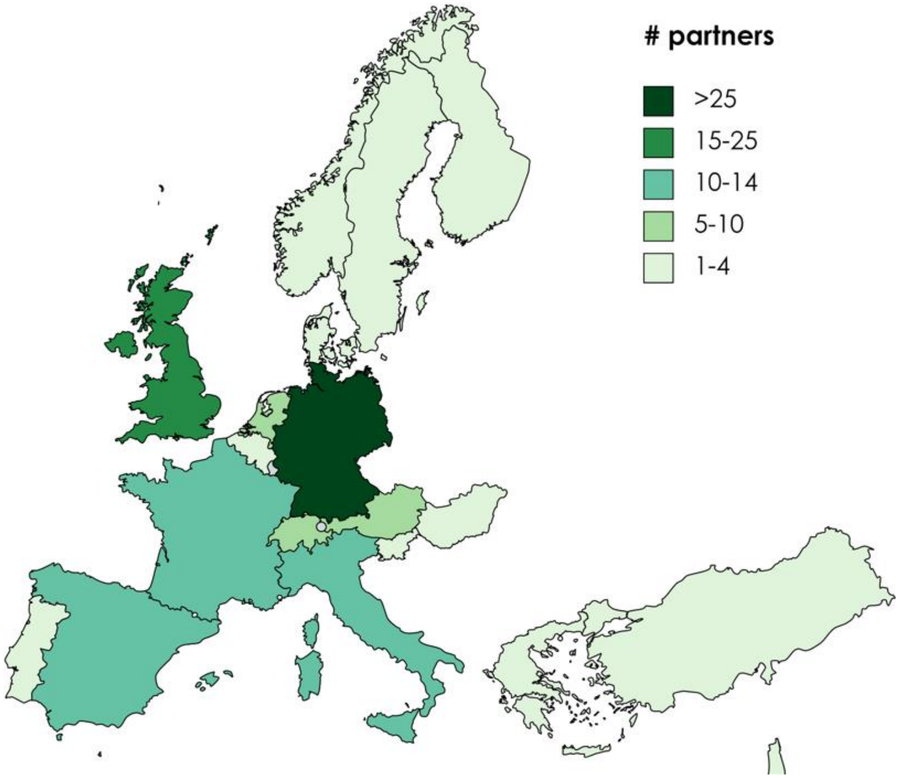
- EU's "Future & Emerging Technology Flagship" project
  - 10-year (2013.10-2023.9), €1B megaproject, ramp-up & 3 main phases
- Goal: advance brain science with computational tools (e.g., AI platforms)
  - Build digital platforms & give grants; public/private access areas to resources within project infrastructure
- Broad areas: study the brain (empirical data, brain simulation, ML); apply the knowledge (medical applications); sharing data, tools & resources (build digital research infrastructure)



HBP overview: geographic distribution of partners

By the end of phase 2 (2020.3.31): 130 partners in 19 countries





Note: Authors' graph. 5 countries only participated in phase 0 are excluded (# partners, €): Canada (1), China (1, 187.5k €),Cyprus (1, 106.2k €), Japan (2), United States (4)

Research Question:

Q: How does the HBP affect the rate and direction of R&D in neuroscience research (publications)?

- Intuitively, productivity can go different ways (an empirical Q):
  - More: “Time to explore new areas and expand the network!”
  - Less: “Time to take more risks and invest in new yet slower areas!”
  - Similar: “Time to change $ source, but I still only have 24 hours/day!”
- Document the progress and development in the HBP (phases 0-2)
- Do researchers actively engaged in the HBP produce more, higher-quality, and/or more interdisciplinary work?
- How does the impact of HBP differ across researchers’ career stage, gender, and sub-fields within neuroscience and computer sciences?

Conceptual Considerations

- Innovation gets harder with the expanding knowledge frontier
  - Need large research teams w/ complementary skills (Jones 2009)
  - More resources and knowledge recombination cross-fields
- But large teams can be inefficient & non-creative
  - Moral hazard with credit sharing (Che & Yoo 2001)
- Most innovation projects are risky, unpredictable, long-term
  - Large spillover effects from public R&D and long-term projects (Williams 2013; Myers & Lanahan 2022),
  - but firms/grants underinvestment in long-term projects (Azoulay et al. 2011; Budish et al. 2015)
- Megaprojects are on the rise in numbers & controversy
  - … often involve infrastructure building, large teams, long term; many underperform (Denicol et al. 2020)

Data: Institutions, individuals, and publications

- **Individuals**: details on 639 active individuals (phases 0-2)
  - HBP involvement (phase active, role), demographic info (gender, affiliation, seniority), research field, highest degree areas, Scopus ID
- **Publications and citations**: Scopus
  - Publication history of individuals, N=39,524 (2008-2022)
- **Core project partners**: details on 173 core project partners by phase 3
  - Department/unit involved, timing of engagement, grants
- **Phase-specific data**: funding amounts, reports, and deliverables
- **Sources**: HBP websites and YouTube channels, deliverables and reports, framework partnership agreements and amendments, Community Research and Development Information Service (CORDIS, European Commission maintained EU research results), HBP PLUS, search on Google Scholar, LinkedIn, and individual/institution websites

HBP partner structure

Empirical Strategy: staggered Diff-in-Diff (DiD)

- Exploit the time-varying access to HBP resources among researchers who ever actively participated in the HBP

- Individuals fixed effects: comparison within the same person over time

$$y_{it} = \beta HBP_{it} + \delta_i + \delta_t + \varepsilon_{it}$$

  - y_{it} : outcome variables at the author-year-level (log +1, numbers)
    - # pubs, # pubs as the first or last authors
    - # pubs in top CS/neuroscience journals
    - By gender, seniority, and topic areas (next slide)
  - HBP_{it}: indicates if individual i has actively participated in HBP by t
  - δ_t & δ_i: individual and year fixed effects
  - Standard errors are clustered at the individual level
  - Apply current methods (e.g., Callaway & Sant'Anna 2021; stacked did wip)
  - Note: given the HBP-focused sample, there are limitations in the variation

- Ongoing: matching-based DiD w doppelgängers (Sosia, Rose & Baruffaldi 2020); include grants in analyses

mDiD

Matching-based DiD

- Employ mDiD methods using propensity score matching to create control group (Hainmueller,2012; Hainmueller & Xu, 2013)
- List of observable characteristics used in sosia-identified Scopus-“doppelgänger” with the following similarity parameters:
  - first_year: the year of the first recorded publication (+/- 5 years)
  - #co-authors, #publications, #citations (up to the year of treatment)
  - Another researcher is *similar* if her characteristics fall within the margin around the original’s characteristics. If all characteristics are similar, and the other researcher is not a co-author, she is deemed a match
  - (default margins: 2 for year of 1st pub, 20% for other three #s)
  - In progress; we then use different methods to restraint the control group (e.g., geo-location, main fields of publication, gender, /random…)
- We search for matches in each 5 years before the treatment (2009-2012) & relaxed the margins to 30% (can relax further if no matches)
- Caveat: very computationally intensive (even on high-performance server) as the search goes through the entire Scopus database

Measuring direction: paper topic classifications

- Publication counts may not capture nuances in novelty/disruptiveness nor direction. Impactful interdisciplinary work may have limited initial citations, then with exponential growth in citations later (Wang et al. 2015)
- Use neural, prompt-based, LLMs (GPT3.5turbo/4) to classify topics

  Significant advantage in few-shot classification tasks (Brown et al. 2020; Chae & Davidson 2023; Clavie et al. 2023; Korinek 2023; Loukas et al. 2023; Stathoulopoulos 2023)

- Provide task prompt & 31 keywords for probabilistic assignment and grouping them into four upper-level topics:
  - Fundamental neuroscience and neurobiology ("neurobiology")
    - e.g., behavioral neuroscience; cellular neuroscience; clinical neuroscience
  - Neurotechnology, simulation, and computational tools ("neurotechnology")
    - e.g., SpiNNaker; brain simulation; cognitive architecture; computational biology
  - Artificial intelligence and robotics ("ai-robotics")
    - e.g., artificial intelligence; machine learning; neurorobotics
  - Clinical applications, treatment, and care management ("clinical research")
    - e.g., brain atlas; clinical trial; management; medical device; neuroethics

GPT-3.5turbo/GPT4 prompt & keywords

- The model assigns weights to each keyword for every abstract-title combination, w/ cumulative weights per article sum up to one
- Ex: ("cognitive neuroscience": 0.4, "virtual reality": 0.3, "neuroethics": 0.15, "behavioral neuroscience": 0.15); we then aggregate to topics

Example of a prompt:

Here's the paper's title and abstract:
Title: Navigating virtual reality by thought: What is it like?
Abstract: We have set up a brain-computer interface (BCI) to be used as an input device to a highly immersive virtual reality CAVE-like system. We have carried out two navigation experiments: three subjects were required to rotate in a virtual bar room by imagining left or right hand movement, and to walk along a single axis in a virtual street by imagining foot or hand movement. In this paper we focus on the subjective experience of navigating virtual reality "by thought," and on the interrelations between BCI and presence.

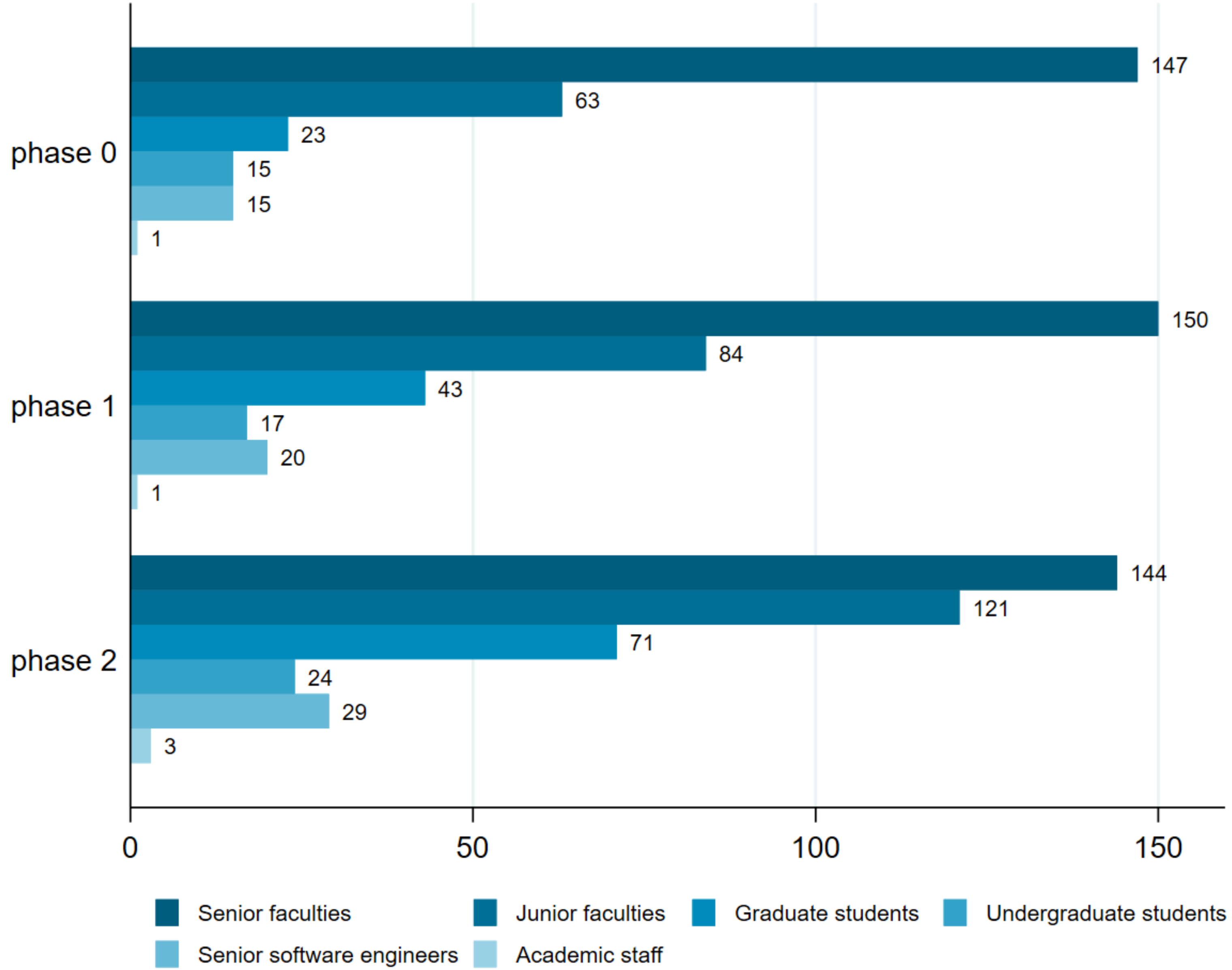
The list of possible topics is:
brain atlas
medical information
neurological treatment
… [omit here for simplicity]

Classify the topic of the paper. Output in the following format with topics and corresponding weights that sum to 1:
{"topic 1": weight 1,
"topic 2": weight 2,
...
"topic n": weight n}

Response to the prompt:

{"artificial intelligence": 0.2,
"cognitive neuroscience": 0.3,
"neuroinformatics": 0.2,
"virtual brain": 0.3}

Descriptive: participants by phase & seniority level

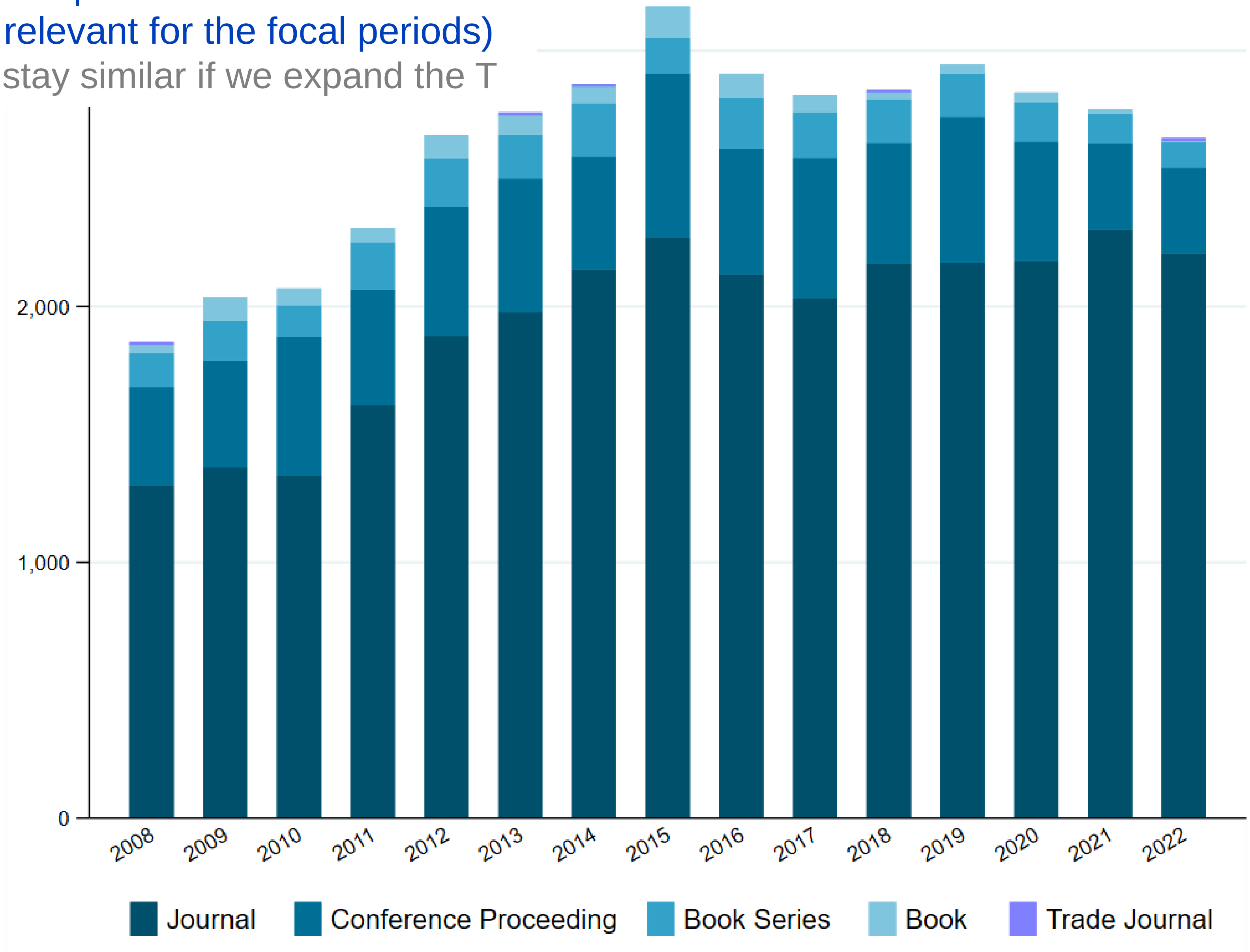


Summary statistics: unique individuals & papers

Panel A: Individual level	Freq.	%	*Panel B: Publication level*	Freq.	%
Total	639	100	Total pubs	39,524	100
Female	129	20.19	*Type of publication*		
Male	511	79.81	Article	24,998	63.25
Active researchers per phase			Conference paper	9,095	23.01
Phase 0	266	41.63	Review	2,107	5.33
Phase 1	318	49.77	Book Chapter	1,087	2.75
Phase 2	393	61.60	Other (e.g., editorial, letter)	2,237	5.66
Countries of affiliation			*Type of journal*		
Phase 0	23		Journal	29,091	73.60
Phase 1	20		Conference proceedings	7,129	18.04
Phase 2	21		Book / book series	3,294	8.33
			Trade journals	10	0.03

Histogram of publications by ever-HBP individuals

Our full sample: 2008-2022
(to stay relevant for the focal periods)
Results stay similar if we expand the T



Results: author-year panel, log DV

- Do researchers actively engaged in the HBP produce more work?

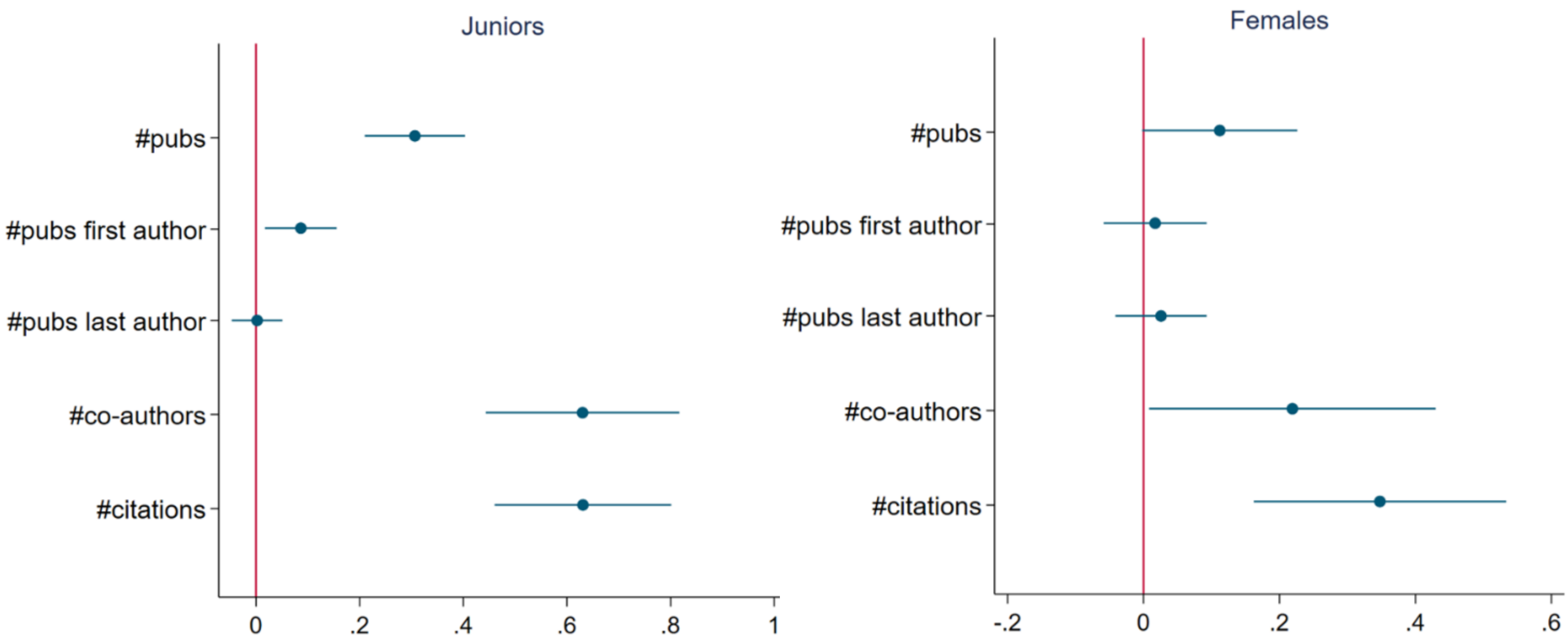
	(1) #pubs	(2) 1st author	(3) last author	(4) #co-authors	(5) #citations
HBP	0.140***	0.0322**	0.0430**	0.294***	0.281***
	(0.0255)	(0.0162)	(0.0178)	(0.0483)	(0.040)
LHS mean	4.1235	0.3528	1.3752	24.5028	271.553
Observations	9,585	9,585	9,585	9,585	9,585
#authors	639	639	639	639	639

Note: Robust standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

- Researchers show a higher productivity and receive more citations after HBP participation (note: w/ individual fixed effects)
- Participation in the HBP expands researchers' networks (# of distinct co-authors per author-year in revealed publications)

Junior/female subsample: author-year, 2008-2022

- Does the HBP's impact differ by researchers' career stage and gender



- Juniors (junior faculties and graduate students) derive great benefit from the HBP research infrastructure and collaboration

- Female researchers experience a significant productivity increase

Results by journal quality and topic areas

- Do researchers actively engaged in the HBP produce higher-quality work after HBP participation?

	(1)	(2)	(3)	(4)
Panel A: Journal Quality (Top neuroscience/CS outlets)				
	Top Neuro	Top CS	CS A*	CS A
HBP	0.0266**	0.00746	-0.00279	0.0125*
	(0.0108)	(0.00820)	(0.00334)	(0.00686)
LHS mean	0.2678	0.1129	0.0371	0.0757
Panel B: Topic Classification				
	Neurobio	Neurotech	AI-Robotics	Clinical
HBP	0.0648***	0.140***	0.0784***	0.0522***
	(0.0220)	(0.0234)	(0.0158)	(0.0196)
LHS mean	2.6266	2.2341	0.8288	1.3723
Observations	9,585	9,585	9,585	9,585
#authors	639	639	639	639

- Higher probability of publishing in top neuroscience journals.
- Increased probability of publishing esp. in topics of neurotechnology.

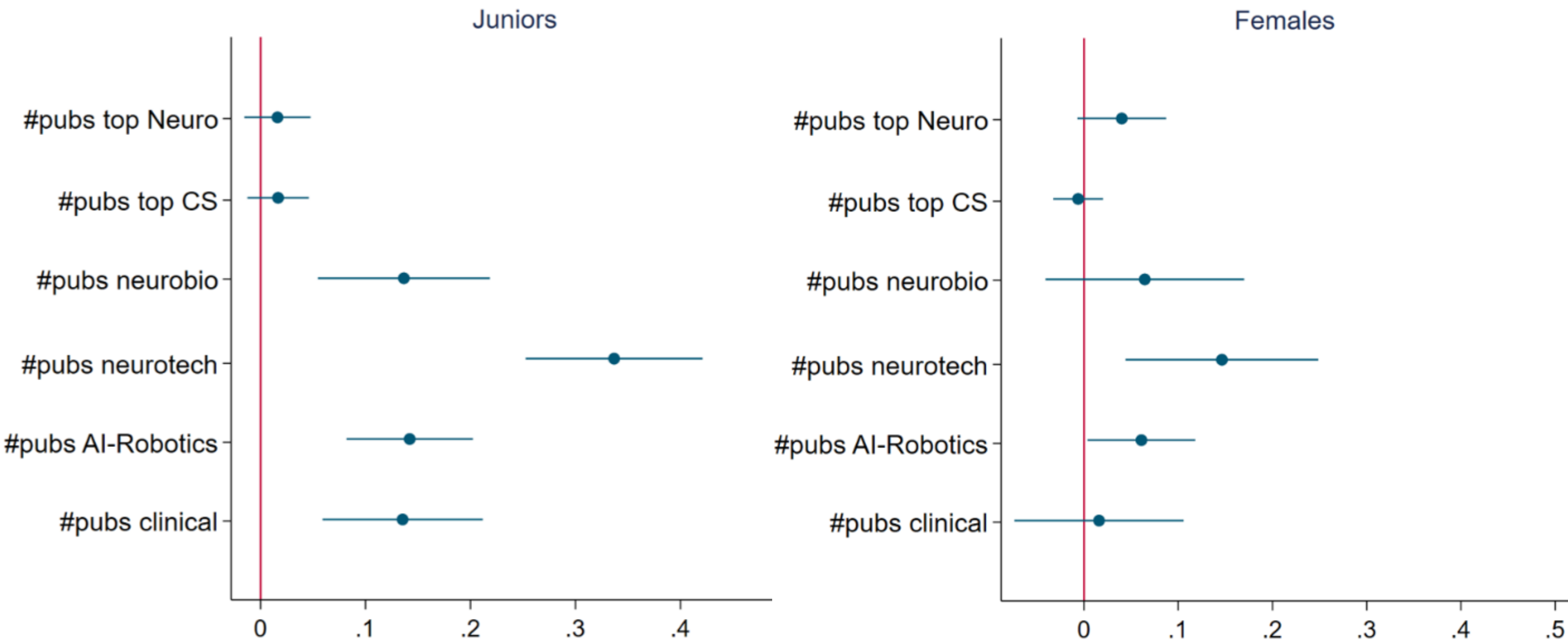
Results: leading topics in top journals

- For researchers actively engaged in the HBP with top-quality work, what are the topics in top neuro journals?

	(1)	(2)	(3)	(4)
Panel D: Topic Classification in Top Neuro Journals				
	Neurobio	Neurotech	AI-Robotics	Clinical
HBP	0.0268**	0.0152**	-0.000559	0.00683
	(0.0106)	(0.00665)	(0.00233)	(0.00675)
LHS mean	0.2476	0.0795	0.0099	0.0552
Observations	9,585	9,585	9,585	9,585
#authors	639	639	639	639

Junior/female quality: author-year, 2008-2022

- Do junior/female researchers produce more top-quality work, or more works in certain topic areas?



- Juniors increases research across topics, especially neurotech areas
- Female yield more top neuro research, & more in neurotech/AI areas

Conclusion & Discussion

The HBP appear to yield positive synergy among AI-neuroscience scholars, and pushed more high-quality interdisciplinary research

- Active involvement in the HBP induces productivity and increased citations, esp. for juniors. Female researchers also boost productivity.
- Researchers have a higher likelihood of publishing in top neuro journals, esp. within the areas of neurobiology and neurotechnology
- Scholars benefit regardless of their affiliated country at the beginning of HBP participation, more for German, Italian, & Belgian-based scholars
- Some evidence that a combination of training and expanded network at the beginning of one's career is most crucial for productivity gains
- Work-in-progress: robustness check using a matching-based DiD w/ doppelgängers; finalizing event studies and various related estimates

Back up slides

Preview of Methods & Results

- Exploit the individual-level active participation in the HBP, with extensive individual fixed effects to partly account for selection
- Benchmark model: diff-in-diff with staggered timing & NLP
- In-progress: matching-based diff-in-diff with various criteria
- Treatment: time-varying access to HBP resources/network among researchers who ever actively participated (vs mDiD control group)

Participation in the HBP leads to …

- (+) Increased individual productivity in publications (esp. juniors)
- (+) Expanded coauthor network, and more citations
- (+) A higher likelihood of publishing in top neuroscience journals
- (+) Increased productivity esp. in neurotech topic areas

Literature and Contribution

- **Institutions and long-term research**: institutions complement market incentives in encouraging valuable innovation, how to better design one

(Arrow 1962; Boudreau 2010; Furman & Stern 2011; Azoulay et al. 2011; Williams 2013; Budish et al. 2015; Murray et al. 2016; Huang et al. 2022; Myers & Lanahan 2022)

- **Science of sciences**: knowledge complementarity and creation within and cross-organizational teams, we focus on a large complex project

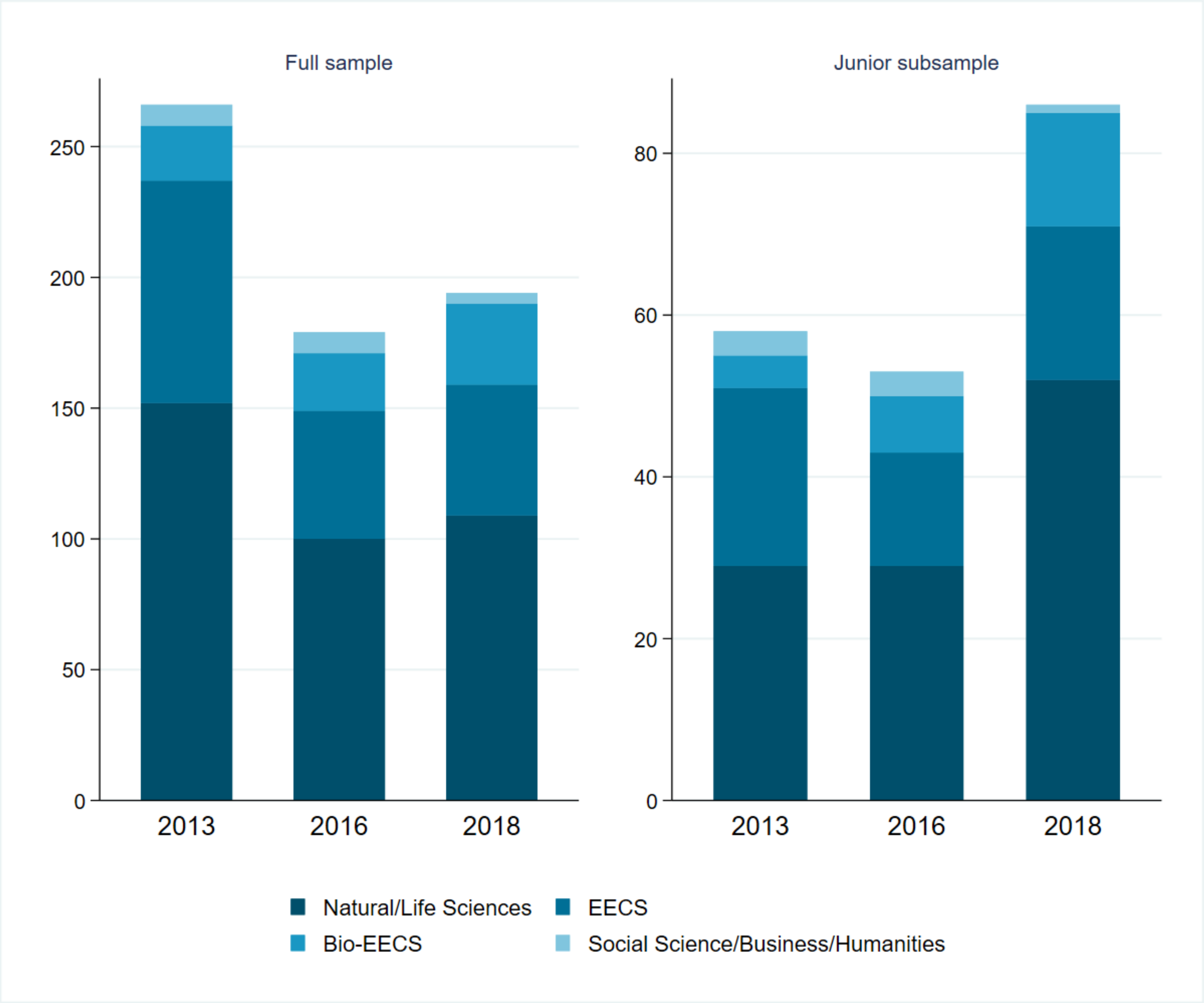
(Nelson & Winter 1985; Cockburn & Henderson 1994; Hagedoorn 2002; Agrawal et al. 2006; Colombo et al. 2006; Melero & Palomeras 2015; Tortoriello et al. 2015)

- **Digitization in health care**: limited evidence and mixed results, mostly on downstream health, and we focus on upstream R&D for future health

(Lou & Wu 2021; Miller & Tucker 2011; McCullough et al. 2016; Agha 2014; Freedman et al. 2017; Goldfarb et al. 2020; Wang 2021)

- **Goal**: 1st empirical analysis on a contemporary long-term science megaproject (in brain-AI) with digital infrastructure and grants

Highest degree fields by individual entering phase



Summary statistics: analytic sample (panel)

Panel C: author-year panel					
	N	Mean	Std. dev.	Min	Max
Panel C1: Publications					
#pubs	9,585	4.12	7.24	0	131
#pubs as first author	9,585	0.35	0.79	0	15
#pubs as last author	9,585	1.38	3.29	0	61
#distinct co-authors	9,585	24.50	66.59	0	1105
#co-authors	9,585	48.93	330.12	0	12731
avg. #co-authors	9,585	5.37	8.96	0	100
#citations	9,585	271.55	677.56	0	11801
Panel C2: Journal Quality					
Top Neuro	9,585	0.27	0.77	0	9
Top CS	9,585	0.11	0.58	0	11
CS A*	9,585	0.04	0.36	0	10
CS A	9,585	0.08	0.43	0	8
Panel C3: Probability of Topic Classification					
Neurobio	9,585	0.24	0.32	0	1
Neurotech	9,585	0.25	0.32	0	1
AI-Robotics	9,585	0.06	0.15	0	1
Clinical	9,585	0.10	0.20	0	1
Panel C4: Topic Classification with the highest probability					
Neurobio	9,585	1.71	3.94	0	46
Neurotech	9,585	1.45	3.57	0	97
AI-Robotics	9,585	0.44	2.02	0	55
Clinical	9,585	0.73	2.27	0	44

Results by field of highest degree

- Training vs. collaboration: HBP interacted w scholars' highest degree

$$y_{it} = \beta HBP_{it} \times field_i + \delta_i + \delta_t + \varepsilon_{it}$$

	(1)	(2)	(3)	(4)	(5)
	#pubs	1st author	last author	#co-authors	#citations
Panel A: Publications and collaborations					
HBPxEECS	0.0623	0.0174	0.0222	0.111	0.260***
	(0.0407)	(0.0217)	(0.0265)	(0.0761)	(0.0852)
HBPxBio-EECS	0.215***	0.0388	0.152***	0.377***	0.272**
	(0.0510)	(0.0344)	(0.0463)	(0.0912)	(0.128)
HBPxNatural/Life Sc.	0.135***	0.00665	0.0252	0.317***	0.170***
	(0.0316)	(0.0186)	(0.0224)	(0.0584)	(0.0602)
Observations	9,585	9,585	9,585	9,585	9,585
#authors	639	639	639	639	639

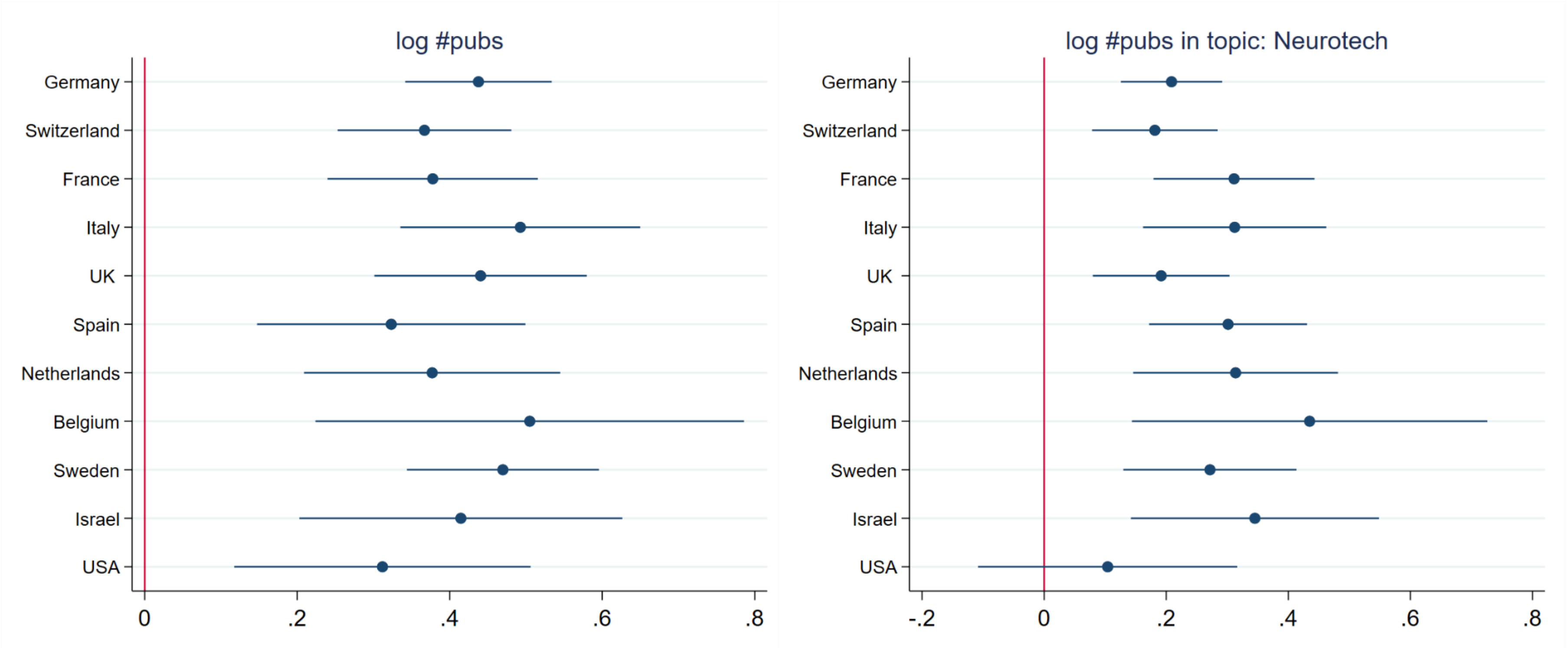
- Note: the base category includes "business, social sciences, and humanities".

- Positive & significant correlation btw the #pubs, #co-authors, #citations & highest degree in natural and life sciences or bio-EECS
- Much stronger results for junior researchers. juniors
- Suggestive evidence that training and network expansion early in one's career are crucial for productivity gains

Log results: full sample – country heterogeneity

- Country-level heterogeneity: HBP interacted w scholars' base country

$$y_{it} = \beta HBP_{it} \times country_i + \delta_i + \delta_t + \varepsilon_{it}$$

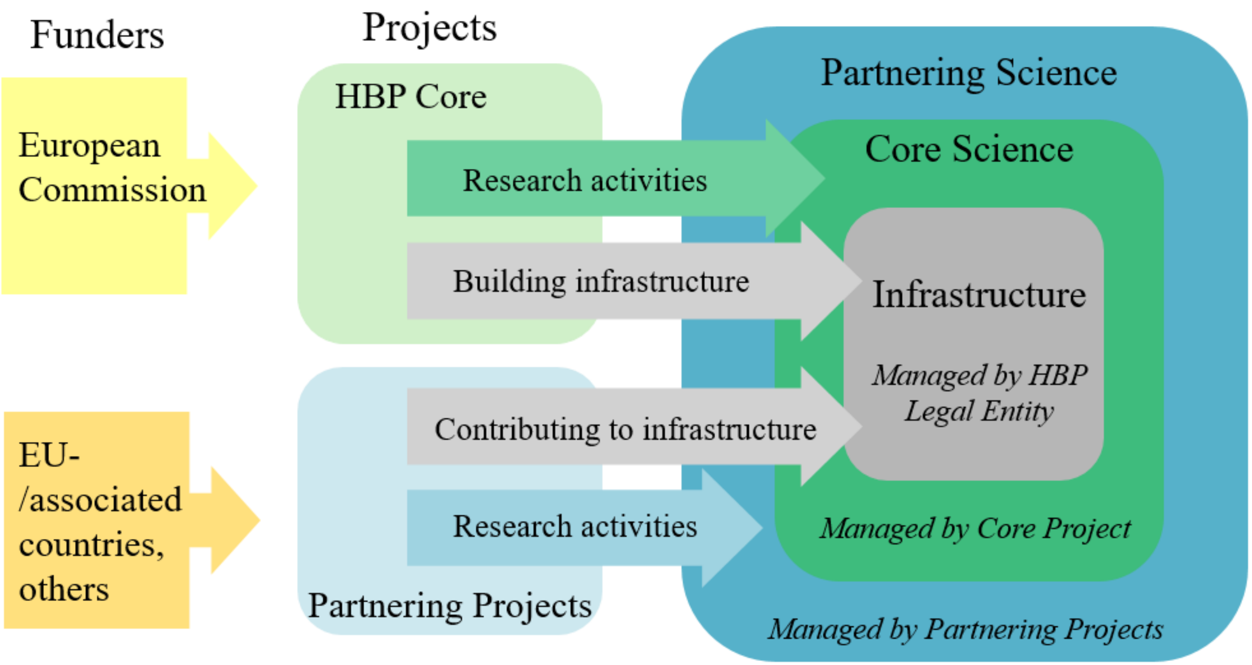


- Note: the base category includes 17 countries w/ <400 obs.

- Positive & significant correlation btw the #pubs & all countries
- Scholars in Germany, Italy, Sweden and Belgium showed significant publication increases, especially in neuroscience & AI-robotics research

HBP structure: core vs partnering project partners

- HBP Core Project Partners:
  - Sign a partnering agreement w/ EC (incl. grant) each phase
  - Are responsible for coordinating & executing the research plan & infrastructure development
- HBP Partnering Projects Partners:
  - Have existing projects w/ own funding
  - Sign an agreement w/ Core Project Partners for selected tasks Form the creative and unrestrained component of the research and infrastructure-building process



back criteria

HBP core project (CP) vs partnering projects

Core project partner application process

- Submit a proposal to the "Call for Expression of Interest"
  (based on Work Plan drawn up by current CP partners for the following phase)
- Eligibility: submit complete, relevant proposals by deadlines
  - Institution of EU-member states/country eligible for EU-grant
  - HBP specific participation criteria (e.g. unique research, unlikely to receive funding from other sources & tight integration across disciplines)
- Evaluation: by the European Commission (EC) & ≥ 3 ext. experts
  - Each evaluation criterion of the call is scored (0-5, 5=excellent)
  - Experts' joint decision is sent to the EC for reviews
  - EC ranks the proposals, # ranked proposals depends on budget
  - Funding decisions are based on ranking: reject vs admit

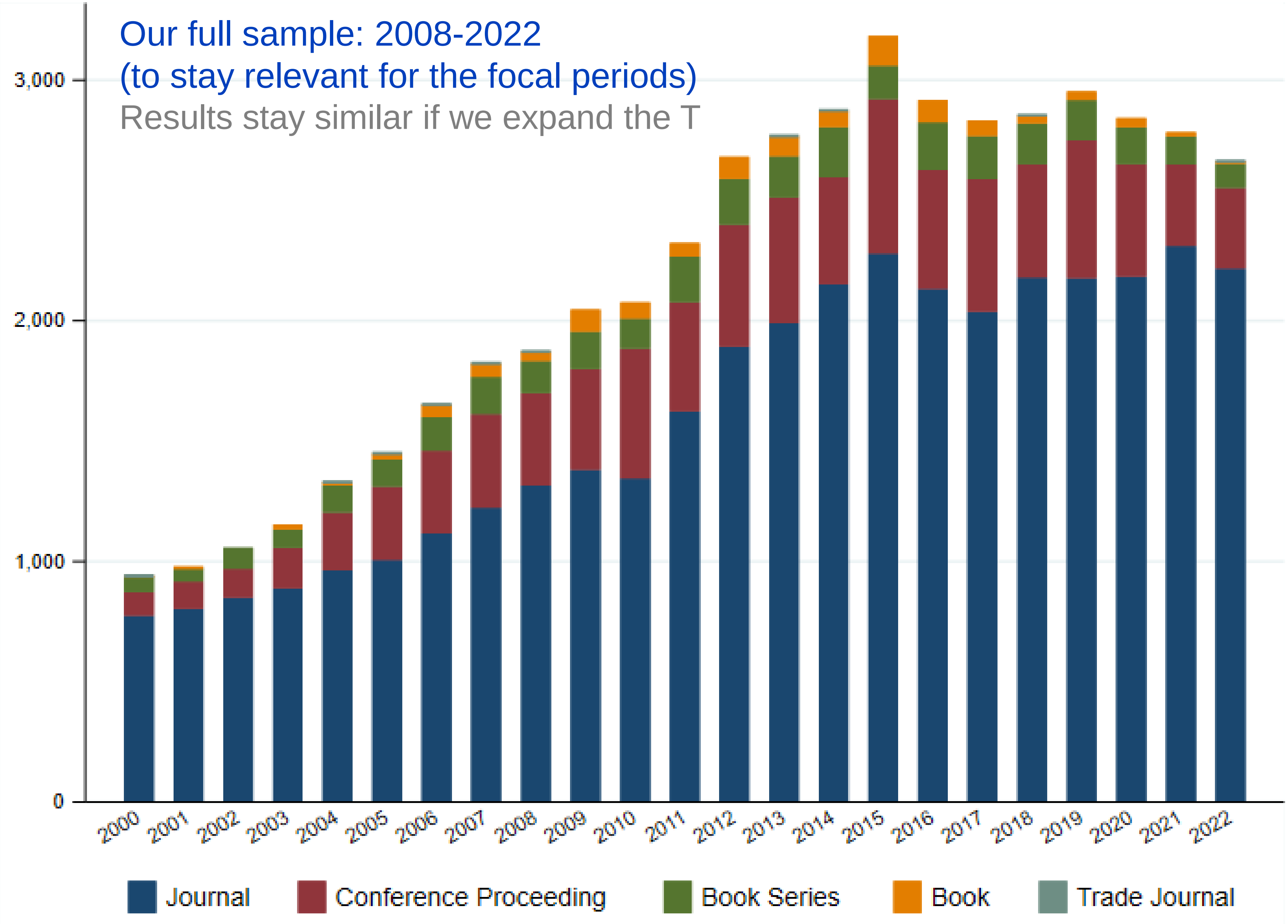
Partnering project partner application process

- Eligibility: can apply at any time
  - Candidates can be nominated by anyone (e.g., EC, HBP CP partners, national/regional funding agencies, private sector, self)
  - Candidate projects need to have their own funding,
  - and should contribute significantly to the Core Project
- Evaluation: by HBP core project (CP) partners
  - CP partners review the proposals and the fit with the respective sub-project
  - Final decision by the HBP Science & Infrastructure Board
  - Memoranda of Understanding btw HBP & partnering project

back

Sources: framework partnership agreements & amendments, EC online portal for funding and tender opportunities

Histogram of publications by ever-HBP individuals



Junior subsample: field of highest degree

	(1)	(2)	(3)	(4)	(5)
	#pubs	1st author	last author	#co-authors	#citations
Panel A: Publications and collaborations					
HBPxEECS	0.166**	0.0511	-0.0189	0.289**	0.450***
	(0.0774)	(0.0465)	(0.0427)	(0.133)	(0.153)
HBPxBio-EECS	0.349***	0.117*	0.00910	0.800***	0.634***
	(0.0972)	(0.0645)	(0.0716)	(0.182)	(0.231)
HBPxNatural/Life Sc.	0.299***	0.0537	-0.000119	0.695***	0.505***
	(0.0646)	(0.0429)	(0.0339)	(0.118)	(0.119)
Constant	0.307***	0.111***	0.0419**	0.553***	1.007***
	(0.0363)	(0.0234)	(0.0207)	(0.0616)	(0.0710)
Observations	2,955	2,955	2,955	2,955	2,955
#authors	197	197	197	197	197

back

Results: full sample country-level heterogeneity