Assignment9_regression

邓会 林定康 丁岩钧 黄永源

2024-01-20

论文基本信息

- **title**： Does Science Advance One Funeral at a Time?
- **author**: Lauren Cohen; Umit Gurun; Danielle Li
- **published**: AUGUST 2019
- **journal**: THE AMERICAN ECONOMIC REVIEW

论文简要介绍

- 本文研究的是科学家死亡能否对他们的所在的研究领域产生重要的影响。
- 作者使用DID方法来验证生物医学研究背景下的“普朗克原理”。
- 研究了452名杰出科学家的早逝如何改变了他们所在子领域的活力,然后与匹配的控制组进行对比。
- 研究表明，这种影响是两方面的，一方面：与早逝的科学家有合作关系的学者发表在该领域的文章会减少，另一方面：新进入者文献数量则会增加。

论文图表复现

小组分工：

- 邓会：table1, table2, figure1, figure3a, markdown
- 林定康：table3,figure2
- 丁岩钧：table6,table7
- 黄永源：table4,table5,figure3b

Table 1—Summary Statistics: Deceased Superstar Scientists

	Mean	SD	Median	Min.	Max.
Year of Birth	1930.16	11.01	1930	1899	1959
Year of Death	1957.63	11.43	1957	1928	1986
Age at Death	1991.13	8.05	1992	1975	2003
Degree year	60.97	9.78	61	34	91
Sudden Death	0.10	0.30	0	0	1
MD Degree	0.40	0.49	0	0	1
PhD Degree	0.49	0.50	0	0	1
MD/PhD Degree	0.11	0.31	0	0	1
Female	0.41	0.49	0	0	1
Nb. of Subfields	6.81	7.31	4	1	57
Career Nb. of Pubs.	138.22	115.70	112	12	1380
Career Nb. of Citations	8341.11	8561.61	5907	120	72122
Career NIH Funding	16637919.28	25441933.30	10899139	0	329968960
Sits on NIH Study Section	0.01	0.08	0	0	1
Career Nb. of Editorials	0.13	1.00	0	0	17

Notes: Sample consists of 452 superstar life scientists who died while still actively engaged in research. See online Appendix A for more details on sample construction.

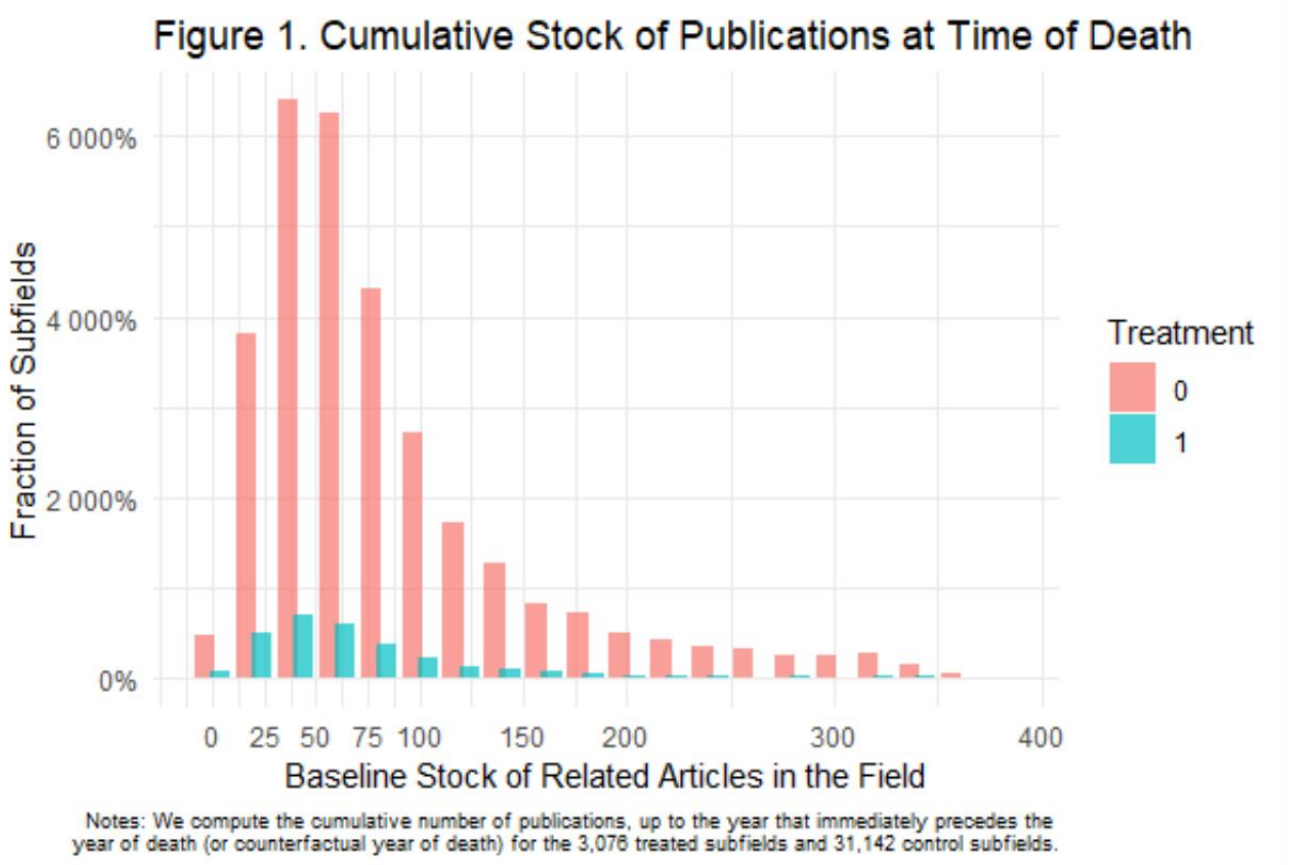
分组描述性统计表格不会弄成分组分别统计，这是在word中把数据分开了。还有三线表横线添加

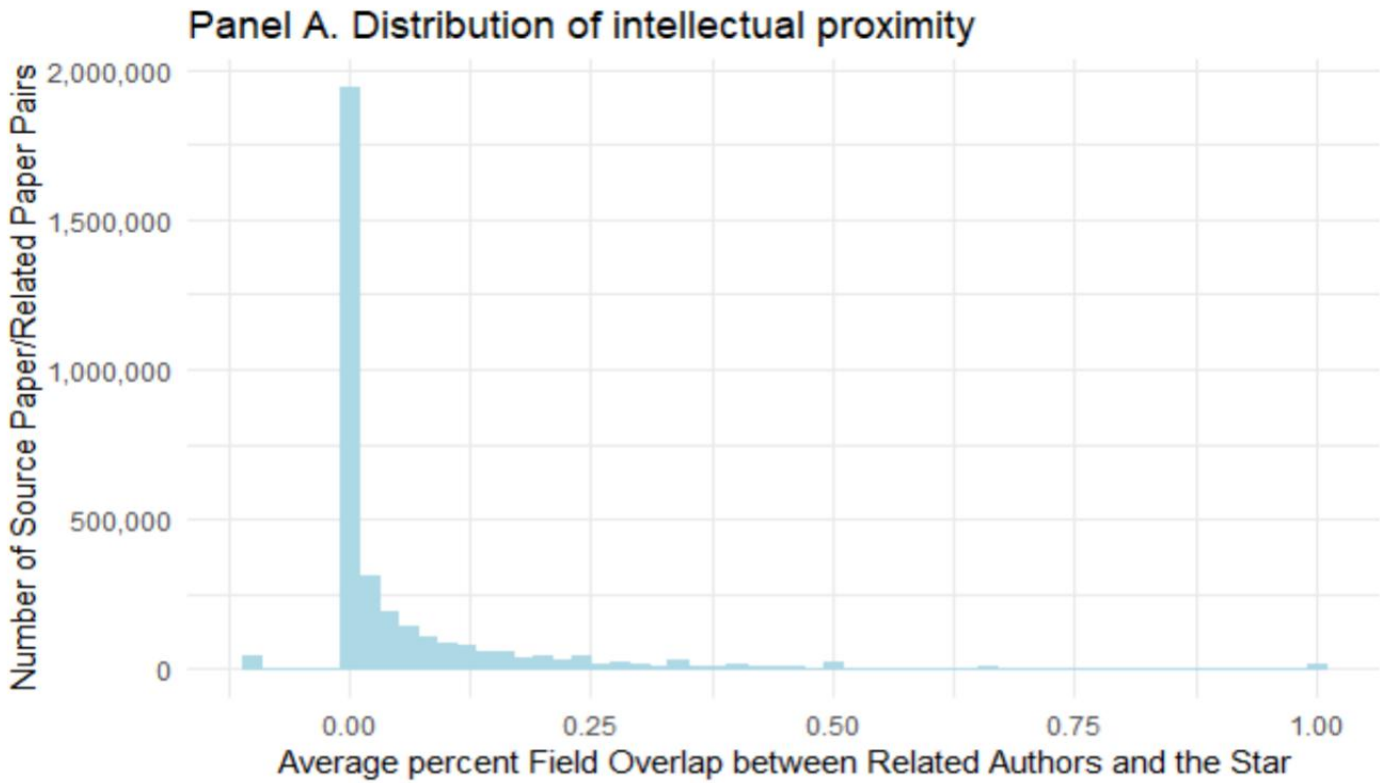
Table 2—Summary Statistics: Control and Treated Subfields at Baseline

variable	mean	median	sd	min	max
Control0	0	0	0	0	0
Baseline Stock of Related Articles in the Field0	85.82	66	66.92	0	384
Baseline Stock of Related Articles in the Field, Non-Collaborators0	76.04	57	62.4	0	381
Baseline Stock of Related Articles in the Field, Collaborators0	9.77	7	10.63	0	125
Investigator Gender0	3.54	3	1.55	1	15
Death Year0	17.72	12	26.41	0	770
Age at Death0	73.88	49	103.59	1	4495
Investigator Cuml. Nb. of Citations0	1965.34	1967	10.33	1926	1991
Investigator Cuml. Nb. of Publications0	0.08	0	0.27	0	1
Source Article Nb. of Authors0	1991.7	1994	7.75	1975	2003
Source Article Long-run Citations0	53.98	54	8.12	34	91
Source Article Citations at Baseline0	137.46	110	102.28	1	1109
Investigator Cuml. NIH Funding at Baseline0	14201603.27	9797410	17821249	0	387558656
Investigator Year of Degree0	11492.22	7834	11866.37	9	157581
Treated1	1	1	1	1	1
Baseline Stock of Related Articles in the Field1	76.28	58	64.05	0	368
Baseline Stock of Related Articles in the Field, Non-Collaborators1	67.75	50.5	59.73	0	357
Baseline Stock of Related Articles in the Field, Collaborators1	8.53	5	9.84	0	86
Investigator Gender1	3.99	4	1.91	1	14
Death Year1	16.69	8	36.33	0	920
Age at Death1	70.43	35	180.53	1	6598
Investigator Cuml. Nb. of Publications1	0.1	0	0.3	0	1
Investigator Cuml. Nb. of Citations1	1960.14	1961	10.9	1928	1986
Source Article Nb. of Authors1	1991.12	1991	7.97	1975	2003
Source Article Long-run Citations1	58.1	58	8.8	34	91
Source Article Citations at Baseline1	169.5	143	118.11	12	1380
Investigator Cuml. NIH Funding at Baseline1	17637725.57	12049690	24873017.73	0	329968960
Investigator Year of Degree1	11579.98	8726	10212.01	120	72122

Notes: The sample consists of subfields for 452 deceased superstar life scientists and their matched control subfields. See online Appendix D for details on the matching procedure. All time-varying covariates are measured in the year of superstar death.

Figure1填充数据有问题，control和treated高度应该是差不多分布的，之后还得调，图片还不太会调整y轴刻度，之前调整，所有刻度都显示在原点。





Notes: Panel A displays the distribution of overlap between the past output of related authors and each star's subfield. For each author on a related article matched to the AAMC Faculty Roster, we create a metric of intellectual proximity by computing the fraction of their publications that belongs to the star's subfield. Slightly more than half of related articles have authors with zero overlap, i.e.,this related article is their first contribution to the star's subfield. 1.24 percent of related articles are authored by new scientists for whom this publication within the subfield is also their first publication overall. Using this information, we aggregate the number of related articles in a particular subfield and in a particular year, e.g., the number of articles in the subfield in year t that have authors above the ninety-fifth percentile in our measure of field overlap.

Figure 2. Effect of Star Scientist Death on Subfield Growth and Decline

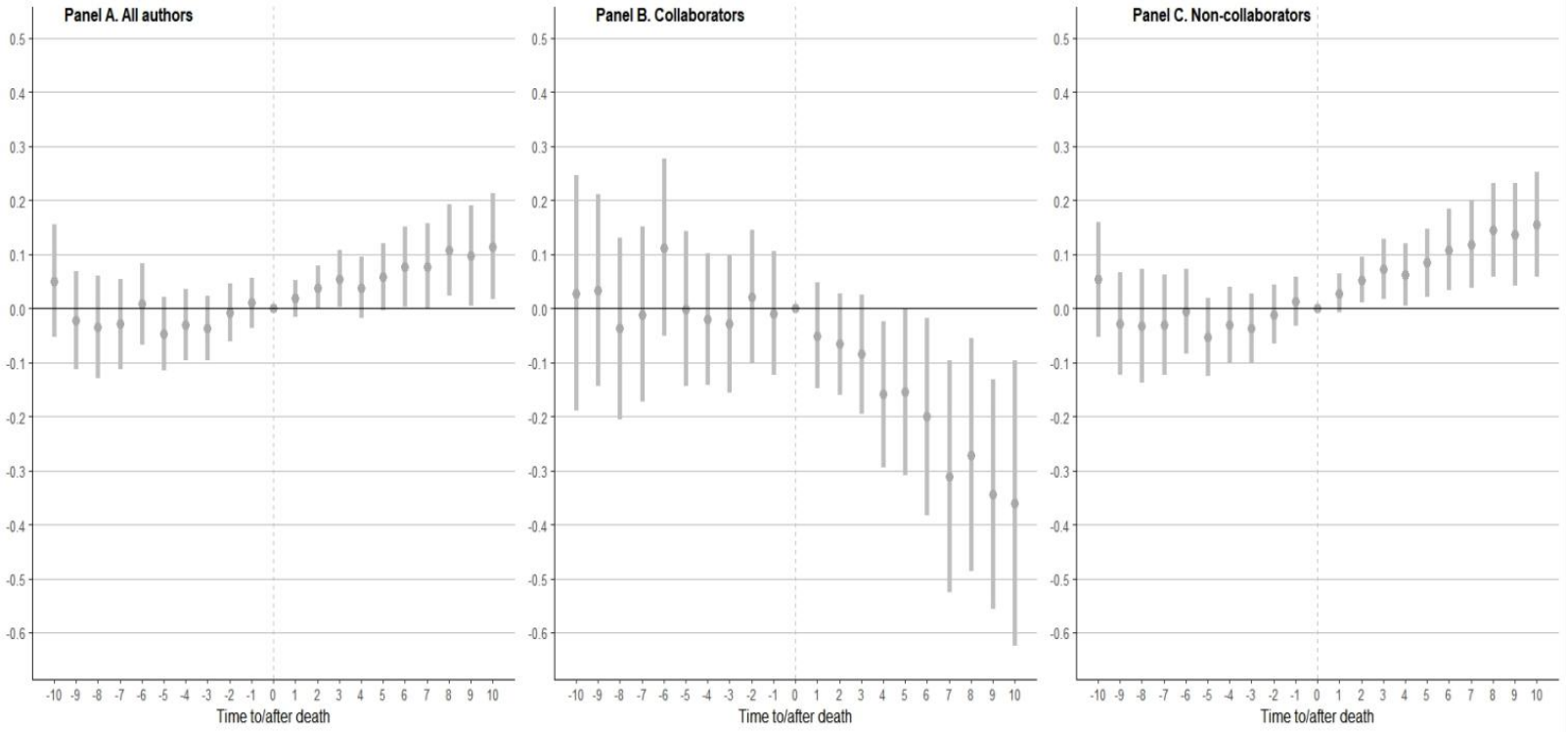


Table 1: Table 3—Effect of Superstar Death on Subfield Entry Rates

	Publication flows			NIH funding flows (number of awards)		
	All authors	Collaborators only	Non-collaborators only	All authors	Collaborators only	Non-collaborators only
	(1)	(2)	(3)	(4)	(5)	(6)
After death	0.051	-0.232	0.082	0.046	-0.265	0.110
	(0.029)	(0.057)	(0.029)	(0.035)	(0.076)	(0.033)
Number of investigators	6,260	6,124	6,260	6,259	6,173	6,257
Number of fields	34218	33096	34218	33912	29163	33806
Number of field-year obs.	1,259,176	1,217,905	1,259,176	1,049,942	902,873	1,046,678
log likelihood	-3,004,997	-799,478	-2,879,836	-1,441,142	-527,470	-1,310,746

Note: Estimates stem from conditional (subfield) fixed effects Poisson specifications. The dependent variable is the total number of publications in a subfield in a particular year (columns 1, 2, and 3), or the total number of NIH grants that acknowledge a publication in a subfield (columns 4, 5, and 6). All models incorporate a full suite of year effects and subfield age effects, as well as a term common to both treated and control subfields that switches from 0 to 1 after the death of the star, to address the concern that age, year, and individual fixed effects may not fully account for trends in subfield entry around the time of death. Exponentiating the coefficients and differencing from one yield numbers interpretable as elasticities. For example, the estimates in column 3 imply that treated subfields see an increase in the number of contributions by non-collaborators after the superstar passes away, a statistically significant $100 \times (\exp[0.082]\ \ 1) = 8.55\%$. Robust standard errors in parentheses, clustered at the level of the star scientist.

Table 6—Breakdown by Star Scientist Characteristics

	Publications		Citations		Funding		Importance to the field	
	Below median	Above median	Below median	Above median	Below median	Above median	Below median	Above median
After death	0.059	0.116	0.036	0.125	0.014	0.162	0.063	0.123
	(0.037)	(0.050)	(0.042)	(0.040)	(0.040)	(0.052)	(0.031)	(0.045)
Number of investigators	2,901	4,836	2,792	4,619	3,048	4,287	5,019	4,493
Number of fields	17210	17008	17328	16890	15731	15487	16985	17233
Number of field-year obs.	632,089	627,087	636,750	622,426	578,277	570,665	625,140	634,036
log likelihood	-1,433,808	-1,443,145	-1,423,746	-1,451,820	-1,319,906	-1,303,402	-1,519,449	-1,312,642

Notes: Estimates stem from conditional (subfield) fixed effects Poisson specifications. The dependent variable is the total number of publications by non-collaborators in a subfield in a particular year. Each pair of columns splits the sample across the median of a particular covariate for the sample of fields (treated and control) in the baseline year. The table examines differences in the extent to which the eminence of the star at death (respectively counterfactual year of death for controls) influences the rate at which non-collaborators enter the field after the star passes away. Eminence is measured through the star's cumulative number of publications, the star's cumulative number of citations garnered up to the year of death, and the star's cumulative amount of NIH funding. We also have a "local" measure of eminence: the star's importance to the field, which is defined as the proportion of articles in the subfield up to the year of death for which the star is an author. All models incorporate a full suite of year effects and subfield age effects, as well as a term common to both treated and control subfields that switches from 0 to 1 after the death of the star. Exponentiating the coefficients and differencing from one yield numbers interpretable as elasticities. For example, the estimate in the second column implies that treated subfields see an increase in the number of contributions by non-collaborators after the superstar passes away, a statistically significant $100 \times (\exp[0.116] - 1) = 12.30\%$. Robust standard errors in parentheses, clustered at the level of the star scientist.

Table 7—The Nature of Entry Barriers

	PMRA-based definition		Citation-based definition		Cliquishness	
	Below median	Above median	Below median	Above median	Below median	Above median
Panel A. Subfield coherence						
After death	0.202	0.067	0.160	0.096	0.129	0.065
	(0.038)	(0.048)	(0.053)	(0.041)	(0.049)	(0.052)
Number of investigators	3,353	3,203	3,418	3,157	2,865	3,552
Number of fields	9062	7828	8722	8159	8044	8820
Number of field-year obs.	334,142	288,284	321,493	300,600	296,704	324,760
log likelihood	-740,575	-690,092	-788,742	-657,865	-719,006	-711,977

	Editorial channel		NIH study section channel		Fraction of subfield NIH funding	
	Below median	Above median	Below median	Above median	Below median	Above median
Panel B. Indirect control through collaborators						
After death	0.163	0.090	0.141	-0.074	0.174	0.084
	(0.055)	(0.049)	(0.042)	(0.097)	(0.051)	(0.051)
Number of investigators	3,452	2,068	4,385	664	3,559	2,525
Number of fields	11110	5780	15338	1552	9863	7027
Number of field-year obs.	410,025	212,399	565,219	57,203	363,690	258,736
log likelihood	-969,457	-473,561	-1,321,692	-129,008	-873,173	-568,556

Notes: Estimates stem from conditional (subfield) fixed effects Poisson specifications. The dependent variable is the total number of publications by non-collaborators in a subfield in a particular year. The sample is limited to the subfields in which the most eminent among the stars were active (specifically, above the median of the "cumulative citations up to the year of death" metric). Each pair of columns splits the sample across the median of a particular covariate for the sample of subfields (treated and control) in the baseline year. For example, the first two columns of panel B compare the magnitude of the treatment effect for stars whose collaborators have written an above-median number of editorials in the five years preceding the superstar's death, versus a below-median number of editorials. All models incorporate a full suite of year effects and subfield age effects, as well as a term common to both treated and control subfields that switches from 0 to 1 after the death of the star. Exponentiating the coefficients and differencing from one yield numbers interpretable as elasticities. For example, the estimates in the first column of panel B imply that treated subfields see an increase in the number of contributions by non-collaborators after the superstar passes away, a statistically significant $100 \times (\exp[0.147] - 1) = 15.84\%$. Robust standard errors in parentheses, clustered at the level of the star scientist.

Table 4—Scientific Impact of Entry

	Vintage-specific long-run citation quantile						
	All pubs	Bottom quartile	2nd quartile	3rd quartile	Btw. 75th and 95th percentile	Btw. 95th and 99th percentile	Above 99th percentile
After death	0.082	-0.028	0.008	0.031	0.125	0.232	0.320
	(0.029)	(0.036)	(0.033)	(0.032)	(0.035)	(0.049)	(0.081)
Number of investigators	6260	6222	6260	6257	6255	6161	5283
Number of fields	34218	33714	34206	34212	34210	33207	21852
Number of field-year obs.	1259176	1240802	1258738	1258954	1258880	1221952	803629
log likelihood	-2879836	-756989	-1208738	-1523133	-1560331	-604625	-186003

Notes: Notes: Estimates stem from conditional (subfield) fixed effects Poisson specifications. The dependent variable is the total number of publications by non-collaborators in a subfield in a particular year, where these publications fall in a particular quantile bin of the long-run, vintage-adjusted citation distribution for the universe of journal articles in PubMed. All models incorporate a full suite of year effects and subfield age effects, as well as a term common to both treated and control subfields that switches from 0 to 1 after the death of the star. Exponentiating the coefficients and differencing from one yield numbers interpretable as elasticities. For example, the estimates in column 1, panel A imply that treated subfields see an increase in the number of contributions by non-collaborators after the superstar passes away, a statistically significant $100 \times (\exp[0.082] \ 1) = 8.55\%$. Robust standard errors in paren- theses, clustered at the level of the star scientist.

Table 5—Entry and Research Direction

	Cardinal measure		Ordinal measure	
	Intellectual proximate articles	Intellectual distant articles	Intellectual proximate articles	Intellectual distant articles
Panel A After death				
After death	0.091	0.028	0.117	-0.024
	(0.030)	(0.035)	(0.028)	(0.037)
Number of investigators	6228	6099	6260	6017
Number of fields	33375	32232	34218	31712
Number of field-year obs.	1228157	1186589	1259176	1167423
log likelihood	-1721061	-1902837	-1996992	-1711295
	In-field versus out-of-field references		Backward citations to the star's bibliome	
	With in-field references	Without in-field references	With references to the star	Without references to the star
Panel B After death				
After death	-0.023	0.128	0.078	0.152
	(0.041)	(0.031)	(0.036)	(0.034)
Number of investigators	6195	6260	6247	6259
Number of fields	32721	34218	34179	34147
Number of field-year obs.	1204315	1259176	1257747	1256576
log likelihood	-859517	-2619762	-2015392	-1860649
	Vintage of cited references		Vintage of 2-way MeSH term combinations	
	Young	Old	Young	Old
Panel C After death				
After death	0.071	-0.010	0.090	0.029
	(0.035)	(0.034)	(0.033)	(0.036)
Number of investigators	6260	6260	6258	6260
Number of fields	34218	34214	34206	34210
Number of field-year obs.	1259176	1259044	1258732	1258906
log likelihood	-2226172	-1708108	-1950365	-1880872

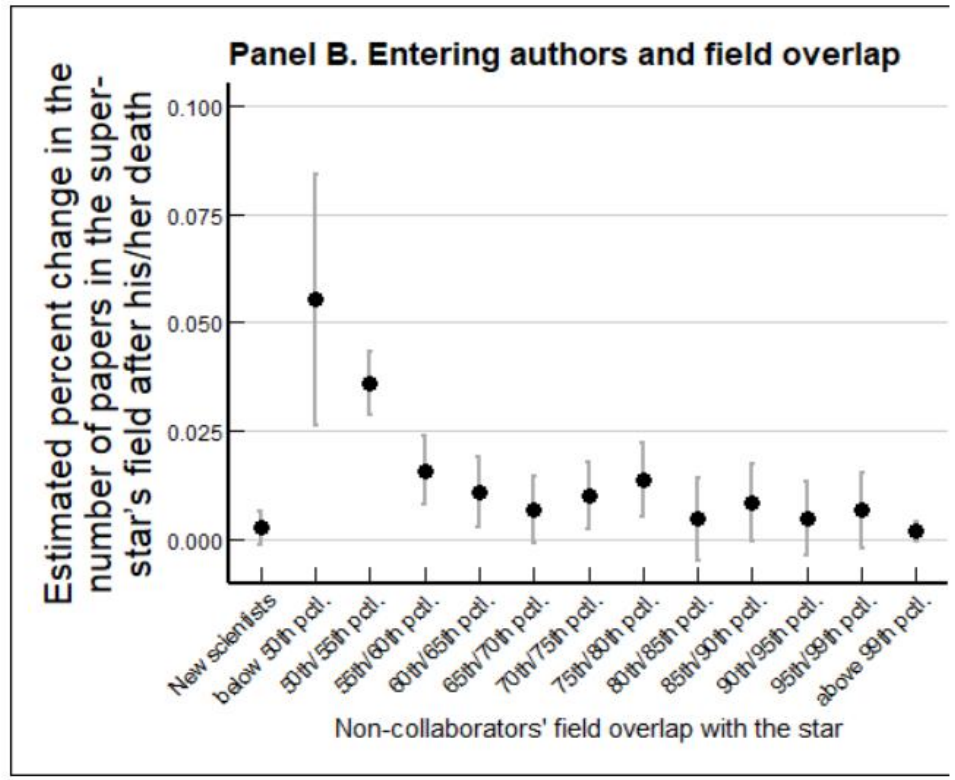


Figure 3. Characteristics of Related Authors: Competitors or Outsiders?