JSCIRES RESEARCH ARTICLE

Migration of scientists to novel areas of biomedical research: Role of article-related productivity

Igor Kissin, Edwin L. Bradley Jr.[1*]

Departments of Anesthesiology, Perioperative and Pain Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, [1]Department of Biostatistics, University of Alabama, Birmingham, Alabama, USA

ABSTRACT

In a previous study (Kissin 2011), frequency of a surname among the general population and the number of articles published by all authors with exactly the same surname were used to estimate surname-based article-related productivity. The main aim of this study was to test the hypothesis that researchers with high productivity are more heavily represented in emerging areas of research than in established areas. As the basis for this study, we used our results on a set of 96 Jewish surname-based groups that have high ratios of actual to expected number of articles (Kissin and Bradley 2013). Seven emerging areas of biomedical research ("angioplasty," "coronary angiography," "genomics," "HIV infections," "meta-analysis," "statins," and "stem cells") were paired with well-established areas. It was found that disproportionally high representation of 96 Jewish surname-based groups in well-established areas (10.2 ± 0.4, mean ± SE) is even higher in emerging areas (18.2 ± 2.6, $P < 0.03$), indicating a disproportional gravitation of these scientists to emerging areas of biomedical research.

Keywords: Biomedical research, emerging research areas, scientific productivity, surnames

INTRODUCTION

In a previous study,[1] frequency of a surname among the general population and the number of articles published in biomedical journals by all authors with exactly the same surname were used to estimate surname-based article-related productivity. In this study, we used this surname-based index to analyze the migration of scientists to novel areas of research. Surnames have recently been used as a proxy for gender, ethnicity, or national origins of scientists. In 2004, Webster analyzed the presence and impact of ethnic minority researchers (seven distinct groups) on science in the UK. He found that in the last 20 years, the number of ethnic-authored papers almost doubled.[2] Lewison and Kundra used surnames to analyze the geographic migration of Indian scientists. They demonstrated that the surname-based analysis can reveal the migration of scientists in all its complexity.[3] A very interesting study was published by Lewison and Markusova on the gender of scientists in Russia. The study was based on the fact that surnames in Russia have gender endings, with "a" denoting a female. The authors found that women had a higher presence in the biological sciences and a very low presence in engineering, mathematics, and physics. The citation scores of female researchers were lower than those of their male counterparts in almost all fields and years.[4] Brookes proposed the concept of "knowledge space" that are multi-dimensional.[5] Liu and Ma demonstrated that the activities of researchers are intensively confined to certain areas in this knowledge space.[6]

The traditional practice and ideology of the academic world gives scientists an opportunity to choose their research areas. This is the most important decision a scientist has to make.[7,8] Carayol and Dalle reviewed various factors that affects the problem of choice, such as

***Address for correspondence:**
E-mail: kissin@zeus.bwh.harvard.edu

Access this article online

Quick Response Code:

Website: www.jscires.org

DOI: 10.4103/2320-0057.135411

the importance of a topic, the difficulty of research, and its audience.[9] The role of energy as a personality trait[10] in the decision to migrate to a novel area of research was behind the direction of this project. We thought that the scientific productivity of a researcher, at least to some extent, depends on the level of his/her energy; and a more-productive researcher is more likely to migrate in an emerging area of research.

The aim of this study was to analyze the migration of scientists to novel, emerging areas of biomedical research. For this analysis, we used 96 Jewish surname-based groups of scientists whose article-related productivity was established previously.[11] We tested the following hypothesis: Researchers with high productivity are more heavily represented in novel, emerging areas of research than in established areas.

MATERIALS AND METHODS

To estimate surname-based article-related productivity, we used the frequency of a surname in the general US population and the number of articles published in biomedical journals by all authors with exactly the same surname. The frequency with which a surname occurred in the US population was determined based on the 2000 Census (www.census.gov/geneology/www/data/2000surnames/index.html, last retrieved 05/20/13). The number of articles published by authors with the same surname was determined using the most comprehensive bibliographic database PubMed, which comprises over 20 million citations of biomedical literature (www.ncbi.nlm.nih.gov/PubMed, last retrieved 05/20/13). All types of articles in all biomedical journals published in English and covered by PubMed were included.

The proportionality of surname-based productivity was determined as a ratio of actual to expected number of articles per thousand persons in a surname group. Ninety-six previously described[11] Jewish surname groups were used in this study: Abramson, Alpert, Auer, Auerbach, Bernstein, Bloch, Brandes, Brodsky, Caplan, Cogan, Cohen, Cohn, Dorfman, Edelman, Edelstein, Epstein, Fleischman, Friedman, Friedlander, Geller, Ginsberg, Ginsburg, Goldberg, Goldfarb, Goldman, Goldstein, Gottschalk, Greenberg, Grossman, Gutman, Halpern, Hellman, Helman, Hirsch, Horowitz, Horwitz, Hurwitz, Jacobsen, Jacobson, Jaffe, Jelinek, Kahn, Kaplan, Katz, Kaufman, Kogan, Kohn, Landau, Lerner, Levin, Levine, Levinson, Levitt, Levy, Lewin, Lieberman, Mandel, Mandell, Margolis, Moskowitz, Pearlman, Posner, Rabinowitz, Rappaport, Resnick, Rosen, Rosenbaum, Rosenberg, Rosenberger, Rosenfeld, Rosenblatt, Rosenblum, Rosenthal, Rubinstein, Schechter, Schorr, Schuler, Schuller, Schulman, Schuster, Schwartz, Segal, Shapiro, Shenk, Shulman, Siegel, Silverman, Silverstein, Steinberg, Wasserman, Weinberg, Weinberger, Weiner, Weintraub, Wexler, and Zuckerman.

The ratio of actual to expected productivity was calculated for three sets of groups. One set represents all 96 surname groups and as a result, R_{96} is the ratio of the actual number of articles (published by authors with surnames belonging to all listed surnames taken together) to the expected number of articles (total number of articles in biomedical journals for a related period of time multiplied by the 96 surname groups share of the total US population). The two other sets of surname groups represent two subdivisions (of the same 96 surnames) divided according to their publication productivity into the higher- and the lower-productivity set. The division was based on the productivity results reported in the previous study (above and below the mean productivity of 38.57 for all 96 surname groups).[11] As a result, R_{42} and R_{54} represent the type of ratios analogous to R_{96}, but related to the higher-productivity set of 42 surname groups and the lower-productivity set of 54 surname groups, respectively.

The performance of these three sets of groups in emerging areas of biomedical research was determined. The selection of these areas was based on two principles: time period (the novel area must have emerged from 1961 to 1990) and the final impact of an emerging area in terms of article publication (total number of articles published in 2008-2012 must have been more than 10,000). The following seven areas were selected: "Angioplasty," "coronary angiography," "genomics," "HIV infections," "meta-analysis," "statins," and "stem cells." The emerging areas of research were paired with established areas, which intended to represent well-established fields within the same category, but much broader than the emerging area. For example, "infections" versus "HIV infections," or "drugs" versus "statins." A paired area was also from the same time interval as an emerging area.

Research productivity in emerging areas was estimated using two approaches. One approach was based on a fixed number of initial publications in an emerging area: the first 1,000 articles (the time periods necessary to collect this

number differed from area to area). The other approach was based on a fixed time interval: 5-year period (starting with a date after the last article of the first 1,000). At least three 5-year periods were used for comparisons for all seven emerging areas.

The significance of differences in research productivity between emerging and paired established areas of research (or between different sets of surname-based groups of authors) was assessed by the following two approaches. First, we calculated the average difference between the emerging areas and their paired established areas and evaluated these differences utilizing a paired *t*-test.[12] The paired *t*-test is a robust statistic that determines whether the average difference between the emerging areas and their paired established areas is a true difference or just due to random chance. The 0.05 level of statistical significance ($P < 0.05$) was used to determine the ability of a ratio to distinguish the emerging areas from their counterparts.

Second, we computed the percentage of times that a ratio for emerging areas exceeded the ratio for paired established areas (or a ratio for one set of groups of authors exceeded the ratio for another set of groups). A 95% confidence interval for each ratio was determined and *P* value was calculated using the nonparametric sign test.[13] Results were declared statistically significant if $P < 0.05$.

RESULTS

The results of the representation of the set of 96 surname-based groups are represented in Tables 1 and 2a-c and provide a comparison between emerging areas of research and paired established areas. The ratios of the actual number of articles to the expected number of articles (96 surname groups, R_{96}) indicate a profound overrepresentation both in emerging and in established areas. The actual number of articles is usually more than 10 times higher than the expected number of articles. However, the degree of overrepresentation is always greater in the emerging areas of research than in the established areas. This difference is apparent in both approaches for the calculation of the ratio (R_{96}), as a share of the first 1,000 of publications [Table 1], or as a share of publications during a 5-year time period [Table 2]. Calculated as a share of the first 1,000 of publications, mean R_{96} in all seven novel areas of research was 17.1 ± 1.6 compared to 11.4 ± 0.5 in the established areas ($P < 0.01$ for the difference). Calculated as a share of publications during the first 5-year period [Table 2a], R_{96} was 18.2 ± 2.6 versus 10.9 ± 0.4 ($P < 0.03$ for the difference). Differences between emerging and established areas persisted during the second and third 5-year periods [Figure 1], but showed a tendency to decrease gradually. The most profound rise of the R_{96} was in the area of "coronary angiography." The complete time course of the changes of R_{96} in this area is illustrated in Figure 2.

When a certain area of research loses its novelty, the degree of disproportionality for the set of 96 groups decreases despite a continuing increase in the total number of publications in that area. Figure 3 illustrates this phenomenon in the area of "angioplasty;" all other emerging areas showed the same type of relationships.

The R_{96} values for emerging areas were statistically higher than those for paired established areas. When they were evaluated with the use of paired *t*-test, the level of significance was always below 0.05 [see the difference columns in Tables 1 and 2a-c]. When statistical significance was estimated

Table 1: Representation of the set of 96 groups (R_{96}*) in emerging areas of research and paired established areas: First 1,000 articles

Emerging area					Paired established area					Difference	
Key words	Total number of articles	Years	96 groups		Key words	Total number of articles	Years	96 groups		R units	%#
			Share of articles	R_{96}*				Share of articles	R_{96}*		
Angioplasty	1007	71-84	85	21.1	Procedure	294,819	71-84	13,960	11.8	9.3	+79
Coronary angiography	1022	61-75	94	22.9	Procedure	103,870	61-75	4567	11.0	11.9	+108
Genomics	1046	87-91	67	16.0	Biology	52,325	87-91	2160	10.3	5.7	+55
HIV infections	1022	82-89	77	18.8	Infections	124,059	82-89	6716	13.5	5.2	+39
Meta-analysis	1006	79-91	49	12.2	Analysis	886,248	79-91	42,225	11.9	0.3	+3
Statins	1006	74-90	70	17.4	Drugs	160,251	74-90	7422	11.6	5.8	+50
Stem cells	1001	61-75	45	11.2	Biology	44,217	61-75	1662	9.5	1.8	+18
Mean±SE				17.1±1.6					11.4±0.5	5.7±1.5**	+50±13

*The relationship of the actual number of articles to expected number of articles (the total number of articles multiplied by a set of groups' share of the total US population), #Compared to R_{96} of the established area, **$P<0.01$. SE=Standard error, HIV=Human immunodeficiency virus

using the percentage of times that a ratio for emerging area exceeded the ratio for its paired established area, the *P* value for the R_{96} difference was < 0.01 ($P = 0.0078$) for each of the four Tables 1 and 2a-c.

A comparison of the representation of the higher-productivity set (42 surname groups, R_{42}) with the lower-productivity set (56 surname groups, R_{56}) in the emerging areas of research is presented in Tables 3 and 4a-c. The ratio of the actual number of articles to the expected number of articles, when calculated as a share of the first 1,000 publications, was 22.8 ± 2.5 for the set of 42 groups and 14 ± 1.4 for the set of 56 groups with $P < 0.002$ for the difference between them [Table 3]. Approximately, the same degree of difference between R_{42} and R_{54} was found when they were determined as a share of publications during the first 5-year periods [Table 4a]. This demonstrates that overrepresentation of the higher-yield productivity set (R_{42}) in the emerging areas of research reached a new maximum – up to 24.8 ± 4.7 times more than expected. Differences in the representations for both sets (42 and 56) of groups during three 5-year periods are presented in Figure 4. They were statistically significant for all three 5-year periods.

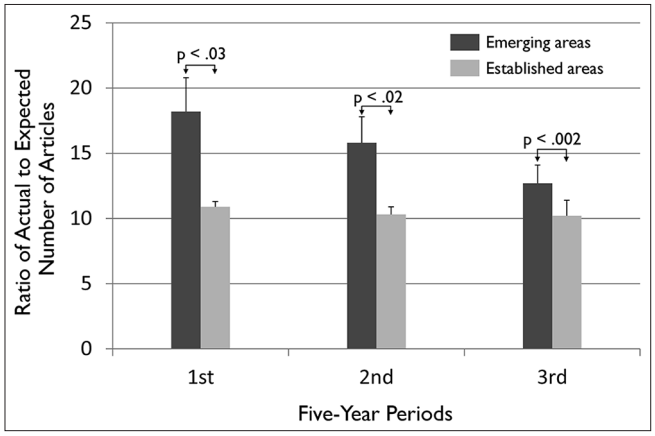

Figure 1: Differences in representation of the set of 96 Jewish surname-based groups (R96) between emerging and established areas. Vertical columns represent the ratio of actual to expected number of articles for the set of 96 surname-based groups, mean ± standard error. Seven emerging areas of research are: "angioplasty," "coronary angiography," "genomics," "HIV infections," "meta-analysis," "statins," and "stem cells." The emerging areas were paired with well-established areas of the same category [but much broader than the respective emerging areas, Table 1]. *P* indicates the level of statistical significance

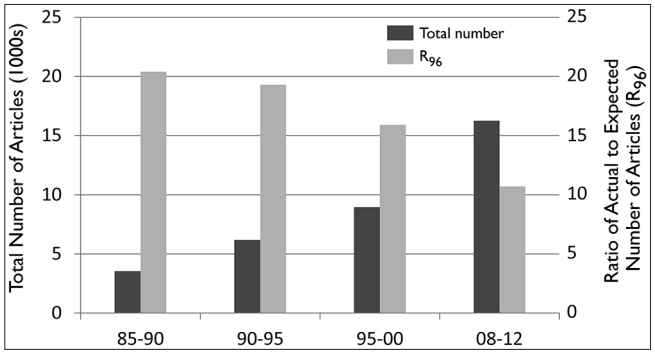

Figure 3: Total number of articles versus the ratio of actual to expected number (R96) during emergence of "angioplasty" area

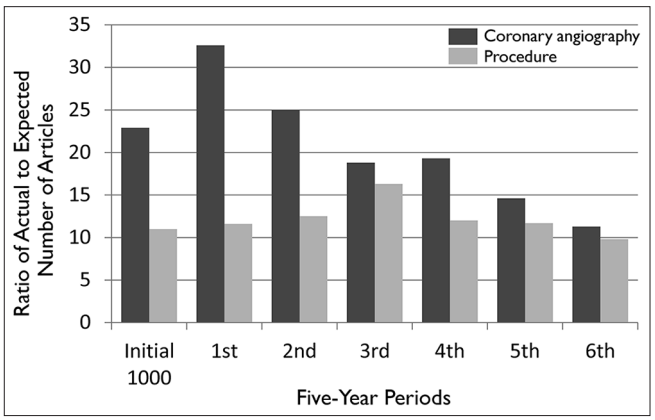

Figure 2: Time-course of the changes in representation of the set of 96 groups during emergence of "coronary angiography" area, 1961-2012. Vertical columns represent the ratio of actual to expected number of articles (R96), "coronary angiography" vs. "procedure" (paired established area). Along the horizontal axis, the first 1,000 articles (articles collected during 1961-1975; then 5-year periods starting with 1975-1980 (1st period)

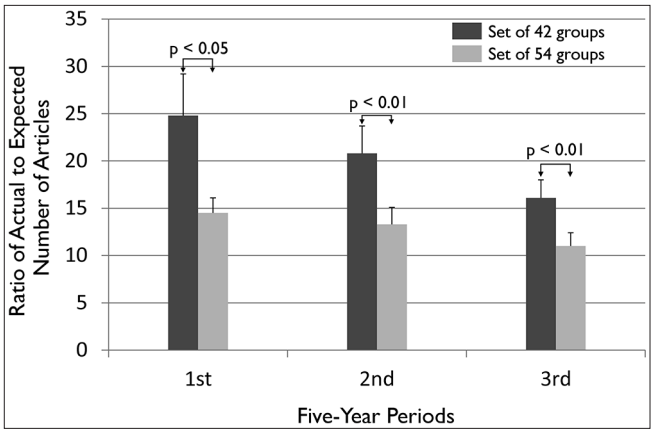

Figure 4: Differences in representation of higher-yield (R42) and lower-yield (R54) sets of groups in the emerging areas of research. Vertical columns represent the ratio of actual to expected number of articles for the set of 42 groups (higher-yield productivity) and the set of 54 groups (lower-yield productivity), mean ± standard error. Emerging areas of research include "angioplasty," "coronary angiography," "genomics," "HIV infections," "meta-analysis," "statins," and "stem cells"

Table 2a: Representation of the set of 96 groups (R_{96}*) in emerging areas of research and paired established areas: First 5-year period

Emerging area					Paired established area					Difference	
Key words	Total number of articles	Years	96 groups		Key words	Total number of articles	Years	96 groups		R units	%#
			Share of articles	R_{96}*				Share of articles	R_{96}*		
Angioplasty	3551	85-90	289	20.4	Procedure	195,928	85-90	9687	12.4	8.0	+65
Coronary angiography	1186	75-80	153	32.6	Procedure	111,255	75-80	5181	11.6	21.0	+181
Genomics	3763	91-96	219	14.6	Biology	91,834	91-96	3678	10.0	4.6	+46
HIV infections	14,306	89-94	843	14.7	Infections	124,903	89-94	6000	12.0	2.7	+23
Meta-analysis	2946	91-96	149	12.6	Analysis	548,297	91-96	23,870	10.9	1.7	+16
Statins	1601	90-95	91	14.2	Drugs	82,150	90-95	3278	10.0	4.2	+42
Stem cells	1795	75-80	132	18.4	Biology	13,255	75-80	511	9.6	8.8	+92
Mean±SE				18.2±2.6					10.2±0.4	7.3±2.5**	+66±21

*The relationship of the actual number of articles to expected number of articles (the total number of articles multiplied by a set of groups' share of the total US population), #Compared to R_{96} of the established area, **$P<0.03$. SE=Standard error, HIV=Human immunodeficiency virus

Table 2b: Representation of the set of 96 groups (R_{96}*) in emerging areas of research and paired established areas: Second 5-year period

Emerging area					Compared wider area					Difference	
Key words	Total number of articles	Years	96 groups		Key words	Total number of articles	Years	96 groups		R units	%#
			Share of articles	R_{96}*				Share of articles	R_{96}*		
Angioplasty	6183	90-95	477	19.3	Procedure	299,014	90-95	14,378	12.0	7.3	+61
Coronary angiography	1658	80-85	165	25.0	Procedure	145,967	80-85	7305	12.5	12.5	+100
Genomics	5673	96-01	250	11.0	Biology	128,213	96-01	4560	8.9	2.1	+24
HIV infections	23,136	94-99	1,388	15.0	Infections	159,723	94-99	6963	10.9	4.1	+38
Meta-analysis	6050	96-01	274	11.3	Analysis	658,338	96-01	25,535	9.7	1.6	+16
Statins	2548	95-00	113	11.1	Drugs	102,517	95-00	3638	8.9	2.2	+25
Stem cells	3872	80-85	273	17.6	Biology	13,370	80-85	504	9.4	8.2	+87
Mean±SE				15.8±2.0					10.3±0.6	5.4±1.5**	+50±1.3

*The relationship of the actual number of articles to expected number of articles (the total number of articles multiplied by a set of groups' share of the total US population), #Compared to R_{96} of the established area, *$P<0.02$. SE=Standard error, HIV=Human immunodeficiency virus

Table 2c: Representation of the set of 96 groups (R_{96}*) in emerging areas of research and paired established areas: Third 5-year period

Emerging area					Compared wider area					Difference	
Key words	Total number of articles	Years	96 groups		Key words	Total number of articles	Years	96 groups		R units	%#
			Share of articles	R_{96}*				Share of articles	R_{96}*		
Angioplasty	8955	95-00	569	15.9	Procedure	478,113	95-00	22,264	11.6	4.3	+37
Coronary angiography	2307	85-90	173	18.8	Procedure	188,038	85-90	9350	16.3	2.5	+15
Genomics	20,206	01-06	707	8.8	Biology	197,707	01-06	5767	7.3	1.5	+21
HIV infections	26,972	99-04	1442	13.4	Infections	198,007	99-04	7831	9.9	3.5	+35
Meta-analysis	11,995	01-06	441	9.2	Analysis	882,291	01-06	3529	8.5	0.7	+8
Statins	7433	00-05	284	9.6	Drugs	147,733	00-05	4456	7.5	2.1	+28
Stem cells	5959	85-90	320	13.4	Biology	46,019	85-90	1913	10.4	3.0	+29
Mean±SD				12.7±1.4					10.2±1.2	2.5±0.5**	+25±4

*The relationship of the actual number of articles to expected number of articles (the total number of articles multiplied by a set of groups' share of the total US population), #Compared to R_{96} of the wider area, **$P<0.002$. SD=Standard deviation, HIV=Human immunodeficiency virus

DISCUSSION

The study findings demonstrated that the general overrepresentation of the set of 96 Jewish surname-based groups in all areas of biomedical research (with the ratio of actual to expected number of articles close to 10) is even greater in emerging areas of research. Calculated as a share of the first 1,000 articles, R_{96} was 17.1 ± 2.6 in the emerging areas of research and 11.4 ± 0.5 in the established areas ($P < 0.01$ for the difference). Calculated as a share of publications

Table 3: Difference in the representation (R*) of two sets of groups (42 vs. 54) during the emergence of novel areas of research: First 1,000 articles

Area of research	Total number of articles in the area	Years	42 groups		54 groups		R_{42}-R_{54} difference	
			Share of articles	R*	Share of articles	R*	R units	%#
Angioplasty	1007	1971-1984	43	27.2	44	17.7	9.5	+54
Coronary angiography	1022	1961-1975	47	29.3	50	20.0	9.3	+47
Genomics	1046	1987-1991	34	20.7	33	12.8	7.9	+62
HIV infections	1022	1982-1989	45	28.1	38	15.1	13.0	+86
Meta-analysis	1006	1979-1991	24	15.2	25	10.0	5.2	+52
Statins	1006	1974-1990	42	26.6	30	12.1	14.5	+120
Stem cells	1001	1961-1975	20	12.7	25	10.2	2.5	+25
Mean±SE				22.8±2.5		14.0±1.4	8.8±1.6**	+64±12

*The relationship of the actual number of articles to expected number of articles (the total number of articles multiplied by a set of groups' share of the total US population), #Compared to R_{54}, **$P<0.002$. SE=Standard error, HIV=Human immunodeficiency virus

Table 4a: Difference in the representation (R*) of two sets of groups (42 vs. 54) during the emergence of novel areas of research: First 5-year period

Area of research	Total number of articles in the area	Years	42 groups		54 groups		R_{42}-R_{54} difference	
			Share of articles	R*	Share of articles	R*	R units	%#
Angioplasty	3551	85-90	126	22.6	178	20.4	2.2	+11
Coronary angiography	1186	75-80	96	51.6	58	20.0	31.6	+158
Genomics	3763	91-96	106	17.9	117	12.6	5.3	+42
HIV infections	14,306	89-94	441	19.6	426	12.1	7.5	+62
Meta-analysis	2946	91-96	85	18.4	66	9.1	9.3	+102
Statins	1601	90-95	42	16.7	55	14.0	2.7	+19
Stem cells	1795	75-80	76	27.0	60	13.6	13.4	+98
Mean±SE				24.8±4.7		14.5±1.6	10.3±3.8**	+70±20

*The relationship of the actual number of articles to expected number of articles (the total number of articles multiplied by a set of groups' share of the total US population), #Compared to R_{54}, **$P<0.04$. SE=Standard error, HIV=Human immunodeficiency virus

Table 4b: Difference in the representation (R*) of two sets of groups (42 vs. 54) during the emergence of novel areas of research: Second 5-year period

Area of research	Total number of articles in the area	Years	42 groups		54 groups		R_{42}-R_{54} difference	
			Share of articles	R*	Share of articles	R*	R units	%#
Angioplasty	6183	90-95	241	24.8	268	17.6	7.2	+41
Coronary angiography	1658	80-85	84	32.3	88	21.6	10.7	+50
Genomics	5637	96-01	123	14.0	129	9.3	4.7	+51
HIV infections	23,136	94-99	733	20.2	693	12.2	8.0	+66
Meta-analysis	6050	96-01	153	16.1	131	8.8	7.3	+83
Statins	2548	95-00	45	11.2	70	11.2	0.0	+0
Stem cells	3872	80-85	164	26.9	120	12.6	14.3	+113
Mean±SE	20.8±2.9					13.3±1.8	7.5±1.7**	+58±13

*The relationship of the actual number of articles to expected number of articles (the total number of articles multiplied by a set of groups' share of the total US population), #Compared to R_{54}, **$P<0.005$. SE=Standard error, HIV=Human immunodeficiency virus

during the first 5-year period, the emerging–established area relationship was almost the same: 18.2 ± 2.6 in emerging areas and 10.2 ± 0.4 in the established areas ($P < 0.03$). Many factors could contribute to the observed overrepresentation in the emerging areas, but the migration of scientists to novel areas of research is most likely the important factor.

The other finding of this study is related to the role of scientists' productivity in these changes. When the 96 surname-based groups were divided into a higher-productivity set (42 groups) and a lower-productivity set (54 groups), there was a clear difference in the degree of representation of these two sets of groups as expressed by the ratio of actual to expected number of articles. Calculated as a share of the first 1,000 articles, R_{42} was 22.8 ± 2.5 and R_{54} - 14 ± 1.4 ($P < 0.002$ for the difference). A similar degree of difference was found when this ratio was calculated as a

Table 4c: Difference in the representation (R*) of two sets of groups (42 vs. 54) during the emergence of novel areas of research: Third 5-year period

Area of research	Total number of articles in the area	Years	42 groups		54 groups		R_{42}-R_{54} difference	
			Share of articles	R*	Share of articles	R*	R units	%#
Angioplasty	8955	95-00	296	21.0	295	13.4	7.6	+57
Coronary angiography	2307	85-90	78	21.5	102	18.0	3.5	+19
Genomics	20,206	01-06	320	10.1	410	8.2	1.9	+23
HIV infections	26,972	99-04	739	17.4	773	11.6	5.8	+50
Meta-analysis	11,995	01-06	217	11.5	233	7.9	3.6	+46
Statins	7433	00-05	128	11.0	158	8.6	2.4	+28
Stem cells	5959	85-90	193	20.5	139	9.5	11.0	+116
Mean±SE				16.1±1.9		11.0±1.4	5.1±1.2**	+48±13

*The relationship of the actual number of articles to expected number of articles (the total number of articles multiplied by a set of groups' share of the total US population), #Compared to R_{54}, **$P<0.006$. SE=Standard error, HIV=Human immunodeficiency virus

share of articles during the first 5-year period – 24.8 ± 4.7 versus 14.5 ± 1.6 ($P < 0.05$). These results indicate that more productive researchers are more heavily represented in emerging biomedical areas.

As a rule, there is a decline in the generation of new knowledge in a given area over time and most of the main discoveries are often found in the first 10 years.[9,14] We found that when a certain area of research loses its novelty, the disproportionality in the representation of more-productive scientists decreases despite a continuing increase in the total number of publications in the area [Figure 3]. This indicates that the overrepresentation of more-productive researchers in the developing area is not governed by a simple increase in the total number of publications in an emerging area.

The disproportionally high representation of the higher-yield productivity set (R_{42}) compared with the lower-yield set (R_{54}) probably reflects the role of some specific personality factors of higher-yield researchers, such as higher levels of energy. Psychological approaches to scientific productivity emphasize the role of several personality traits in enhancing research output, such as the ability to compete, motivation, and especially the ability to take risks (risk-taking ability is commonly expected in a person moving to a new area).[7,15-17]

The mobility of scientists in the knowledge space is very limited. Liu and Ma[6] demonstrated that the activity of researchers is confined to a certain area of research; usually it is limited to changes within the same field of studies. One well-known example of the migration of a scientist to a distinctly different area of studies is that of Ivan Pavlov, who received the Nobel Prize (1904) for his studies on the physiology of digestion, but is now best known for his studies on "conditioning" as an autonomic form of learning. Over approximately 10 years, he changed his research focus from physiology of circulation to physiology of digestion and then to comparative psychology. Herbert Simon also performed research in dramatically different areas. He had important achievements in economics, computer sciences, political sciences, and psychology.

Changes among closely related areas of research are relatively frequent. When an advance in sciences results in a novel development, the decision to rapidly move into this area requires certain personality traits in addition to the regular attributes of a researcher. We suggest that the higher energy levels, reflected by high publication productivity might be one of those traits: moving to a new area always requires more energy. In addition, migration to a new area of research is always associated with greater risk. Therefore, risk-taking ability might be another required trait.

It is of interest that the degree of overrepresentation in the novel areas of research for the set of 42 surname groups was 2.5-fold higher than that with all 96 surname groups for the established areas, 24.8 ± 4.7 versus 10.2 ± 0.4 [Tables 4a and 2a]. The drive to migrate to a novel area of research is probably the strongest in the most productive researchers.

The ratio of actual to expected number of articles for the set of 96 surname-based groups in the established biomedical areas was close to 10 [Tables 1 and 2a-c]. The previously reported ratio related to biomedical articles for the same set of Jewish surname-based groups (R_{96}) was much higher – close to 23.[11] The difference is due to the fact that in the previous study, only articles with abstracts published in 40 of the top biomedical journals

were included. In this study, all types of the articles in all biomedical journals published in English and covered by PubMed were included. Another publication on productivity of nine Jewish surname-based groups (based on the same criteria) reported the ratio of the actual number to expected number of articles as 9.7.[1] This is actually the same value as that observed in the present study in established areas of research.

This study has the following limitation. Some authors related to the indicated above surname-based group may not be US scientists. However, their number is relatively small, and in addition, even if the number of noncitizens reached 5% or 10%, the degree of the observed overrepresentation is large enough (10-25 times) that the role of this factor should be insignificant.

In conclusion, researchers with high productivity are more heavily represented in emerging areas of research than in established areas, indicating their gravitation to novel ideas.

REFERENCES

1. Kissin I. A surname-based bibliometric indicator: Publications in biomedical journals. Scientometrics 2011;89:273-80.
2. Webster BM. Bibliometric analysis of presence and impact of ethnic minority researchers on science in the UK. Res Eval 2004;13:69-76.
3. Lewison G, Kundra R. The internal migration of Indian scientists, 1981-2003, from an analysis of surnames. Scientometrics 2008;75:21-35.
4. Lewison G, Markusova V. Female researchers in Russia: Have they become more visible? Scientometrics 2011;89:139-52.
5. Brookes BC. The foundations of information science IV. J Inf Sci 1981;3:3-12.
6. Liu X, Ma F. Transfer and distribution of knowledge creation activities of bio-scientists in knowledge space. Scientometrics 2013;95:299-310.
7. Merton RK. Priorities in scientific discovery: A chapter in the sociology of science. Am Soc Rev 1957;22:635-59.
8. Ziman JM. The problem of "problem choice." Minerva 1987;25:92-106.
9. Carayol N, Dalle JM. Sequential problem choice and reward system in open science. Struct Change Econ Dyn 2007;18:167-91.
10. Goldberg LR. An alternative "description of personality": The big-five factor structure. J Pers Soc Psychol 1990;59:1216-29.
11. Kissin I, Bradley EL Jr. A surname-based patent-related indicator: the contribution of Jewish inventors to US patents. Scientometrics 2013;97:357-68.
12. Brownlee KA. Statistical Theory and Methodology in Science and Engineering. 2nd ed. New York: Wiley; 1965.
13. Daniel WW. Applied Nonparametric Statistics. Boston: Houghton Milfflin Company; 1978.
14. Crane D. Invisible Colleges: Diffusion of Knowledge in Scientific Communities. Chicago: University of Chicago Press; 1972.
15. Vroom VH. Work and Motivation. Oxford, England: Wiley; 1964.
16. Hagstrom WO. Competition in science. Am Soc Rev 1974;39:1-18.
17. Blackburn RT. Tenure: Aspect of Job Security on the Changing Campus. Atlanta, GA: Southern Regional Education Board; 1972.

How to cite this article: Kissin I, Bradley EL. Migration of scientists to novel areas of biomedical research: Role of article-related productivity. J Sci Res 2013;2:206-13.

Source of Support: Nil, **Conflict of Interest:** None declared

Staying in touch with the journal

1) Table of Contents (TOC) email alert
Receive an email alert containing the TOC when a new complete issue of the journal is made available online. To register for TOC alerts go to **www.jscires.org/signup.asp.**

2) RSS feeds
Really Simple Syndication (RSS) helps you to get alerts on new publication right on your desktop without going to the journal's website. You need a software (e.g. RSSReader, Feed Demon, FeedReader, My Yahoo!, NewsGator and NewzCrawler) to get advantage of this tool. RSS feeds can also be read through FireFox or Microsoft Outlook 2007. Once any of these small (and mostly free) software is installed, add **www.jscires.org/rssfeed.asp** as one of the feeds.