

DOI:

КОРРЕЛЯЦИЯ НАУКОМЕТРИЧЕСКИХ ПОКАЗАТЕЛЕЙ ИЗ РИНЦ С ЦИТИРОВАНИЕМ ПО БАЗЕ WEB OF SCIENCE

Салтыков С.А.

Институт проблем управления им. В.А. Трапезникова РАН,
Россия, г. Москва, ул. Профсоюзная д.65
ssaltykov@mail.ru

Аннотация: В данной работе проанализирована связь между наукометрическими показателями из РИНЦ и числом цитирований по Web of Science (WoS) для различных групп исследователей. Показано, что для исследователей «среднего эшелона» наибольшей корреляцией с числом цитирований по WoS из показателей в РИНЦ обладает средний импакт-фактор журналов, в которых опубликованы работы. Это аргумент в пользу того, что таким исследователям стоит ориентироваться на качество работ в большей степени, чем на количество, если их цель — высокая цитируемость по WoS. Также продемонстрировано, что для таких исследователей число статей в российских журналах (в том числе из списка ВАК) имеет отрицательную корреляцию с числом цитирований по WoS. Это можно интерпретировать так: начинающим исследователям и исследователям «среднего эшелона» нецелесообразно ориентироваться на чрезмерное увеличение числа публикаций в российских журналах, не входящих в ядро РИНЦ, так как это может увести в сторону от создания работ, которые могут получить мировое признание.

Ключевые слова: РИНЦ, Web of Science, наукометрия.

Введение

Если у исследователя высокие показатели цитирования по базе Web of Science (WoS), то это почти точно означает, что у этого исследователя есть международное признание. Однако, далеко не у всех российских ученых есть работы в WoS. Но даже по тем, у кого есть работы в WoS, информацию поглядеть нелегко. Поэтому для непосредственной оценки большинства российских ученых (особенно начинающих, а также гуманитариев) эта база подходит не полностью.

С другой стороны, в базе «Российский индекс научного цитирования» (РИНЦ) есть практически все российские ученые, даже самые начинающие. Найти любого исследователя по ФИО легко, также легко учитывать не только отдельных ученых, но и целые организации и структурные подразделения. Данные по самым различным срезам из РИНЦ можно получать довольно системно. Но только вот что значат эти данные из РИНЦ для конечной цели — создания признанных в мире, цитируемых мировым научным сообществом работ? Какие именно показатели из РИНЦ свидетельствуют о том, что у автора много цитирований в базе WoS, а какие показатели мало с этим связаны?

Таким образом, как совместить преимущества РИНЦ (полнота данных о российских исследователях) и WoS (высокоцитируемые работы в WoS — эквивалент международного признания)? Следуя логике экспертно-наукометрического подхода [1], можно методами машинного обучения построить модель, которая будет предсказывать для данного автора число ссылок из WoS по данным из РИНЦ. Но для начала, возможно, стоит посмотреть, какие показатели из РИНЦ больше скоррелированы с числом цитирований из WoS, а какие меньше.

Цель работы — на небольшом наборе данных из РИНЦ получить статистически значимые результаты относительно того, для каких подмножеств российских исследователей какие показатели из РИНЦ в наибольшей степени коррелируют с числом ссылок по WoS. В дальнейших работах стоит увеличить анализируемый набор данных.

1 Методы исследования и анализируемые данные

Я попытался выкачать из РИНЦ исследователей с идентификационными номерами с 1 по 1000. Удалось выкачать 31 показатель по 997 исследователям. У 153 из них в их анкете в РИНЦ была проставлена ссылка на их показатели в WoS. Для них были выкачаны данные из WoS — сколько ссылок на работы данного автора есть в WoS. Далее будут анализироваться те исследователи, у которых есть хотя бы одно цитирование по данным WoS. Таких исследователей в выборке 101. Таким образом, для них у нас есть число цитирований по WoS и полный набор показателей в РИНЦ. Посмотрим, как значения показателей из РИНЦ коррелируют с числом ссылок по WoS. Описанный датасет, содержащий 977 записей, выложен на github [2].

Исследование построено следующим образом. Рассчитаем ранговую корреляцию Спирмена между значениями каждого из 31 показателей в РИНЦ и числом ссылок из WoS для всего набора из 101 исследователей. Статистически значимыми будем считать корреляции с р-критерием меньше 0,05.

Затем так же проанализируем два подмножества этого набора отдельно — исследователей среднего эшелона и научных «корифеев» и посмотрим, буду ли на подмножествах набора такие же выводы, что и на всем наборе целиком.

В табл. 1, представленной ниже, кроме прочего, перечислены анализируемые показатели, пронумерованные от №0 до №30. Проанализируем сами эти показатели. Существенная доля показателей представляет собой или число статей данного исследователя, или число цитирований работ данного исследователя, или индекс Хирша, рассчитанные по различным совокупностям научных работ. Такими совокупностями научных работ могут быть как базы публикаций — РИНЦ, перечень ВАК, ядро РИНЦ — так и научные работы, отобранные по определенным критериям. Например, работы без учета монографий или только зарубежные журналы. Большинство показателей являются, можно сказать, «экстенсивными»: чем дольше исследователь пишет с одним и тем же качеством, тем больше оказывается этот показатель. Но есть и «интенсивные» показатели, например, №12 «Среднее число цитирований в расчете на одну публикацию», а также показатели №29 и №30, завязанные на импакт-фактор журналов.

Кроме того, стоит отметить, что среди показателей есть такой, который кажется почти «тавтологическим» для цели предсказания числа цитирований по WoS. Это показатель №24 «Число цитирований из зарубежных журналов», он содержит в себе уже цитирования по WoS, но, однако, и другие цитирования тоже. Если окажется, что в нем доля цитирований по WoS очень высока, то он, конечно же, будет лучшим предиктором (так как будет почти предсказывать сам себя). Поэтому если окажется так, что он будет лидировать, интересно будет, какой показатель идет за ним, какой менее «тавтологический» показатель хорошо предсказывает цитирования по WoS.

Также стоит помнить, что база «ядро РИНЦ» содержит в себе работы из Russian Science Citation Index (RSCI) [3], а также из Scopus и WoS.

2 Анализ всего имеющего множества исследователей

Опишем исследуемое множество. В выборку попадают исследователи, у которых в профиле в базе РИНЦ проставлена ссылка на их профиль в WoS, и на работы исследователя есть хотя бы одна ссылка по базе WoS. Из всех исходно анализируемых 977 исследователей этим критериям удовлетворяют только 101. То есть $101 / 977 = 10\%$ исследователей.

Для каждого из показателей посчитаем ранговую корреляция Спирмена с числом цитирований по WoS для данного исследователя, а также статистическую значимость этой корреляции. Полученные данные сведем в таблицу, затем отобразим на рис. 1.

Таблица 1. Показатели в РИНЦ всех анализируемых исследователей

№	Наименование показателя в РИНЦ	Корреляция	p-критерий
0	Число публикаций на elibrary.ru	0.538091	6.506214e-09
1	Число публикаций в РИНЦ	0.547581	3.105668e-09
2	Число публикаций, входящих в ядро РИНЦ	0.690272	1.402123e-15
3	Число цитирований из публикаций на elibrary.ru	0.766964	8.720008e-21
4	Число цитирований из публикаций, входящих в РИНЦ	0.781965	4.827543e-22
5	Число цитирований из публикаций, входящих в ядро РИНЦ	0.870563	2.960983e-32
6	Индекс Хирша по всем публикациям на elibrary.ru	0.786342	1.986902e-22
7	Индекс Хирша по публикациям в РИНЦ	0.810966	8.853414e-25
8	Индекс Хирша по ядру РИНЦ	0.906759	6.529027e-39
9	Число публикаций, процитировавших работы автора	0.753041	1.060774e-19
10	Число ссылок на самую цитируемую публикацию	0.662864	4.315415e-14
11	Число публикаций автора, процитированных хотя бы один раз	0.618802	5.331535e-12
12	Среднее число цитирований в расчете на одну публикацию	0.762427	2.005677e-20
13	Индекс Хирша без учета самоцитирований	0.800955	8.759725e-24
14	Индекс Хирша с учетом только статей в журналах	0.887737	3.975188e-35
15	Год первой публикации	-0.249554	1.184517e-02
16	Число самоцитирований	0.591684	7.236632e-11
17	Число цитирований соавторами	0.718120	2.859078e-17
18	Число соавторов	0.513335	4.036486e-08
19	Число статей в зарубежных журналах	0.848185	4.485363e-29
20	Число статей в российских журналах	0.311978	1.493315e-03
21	Число статей в российских журналах из перечня ВАК	0.341203	4.798345e-04
22	Число статей в российских переводных журналах	0.433844	5.833974e-06

№	Наименование показателя в РИНЦ	Корреляция	p-критерий
23	Число статей в журналах с ненулевым импакт-фактором	0.636478	8.471750e-13
24	Число цитирований из зарубежных журналов	0.913677	1.706968e-40
25	Число цитирований из российских журналов	0.409706	2.094528e-05
26	Число цитирований из российских журналов из перечня ВАК	0.429965	7.211893e-06
27	Число цитирований из российских переводных журналов	0.614035	8.586513e-12
28	Число цитирований из журналов с ненулевым импакт-фактором	0.822426	5.399997e-26
29	Средневзвешенный импакт-фактор журналов, в которых были опубликованы статьи	0.800582	9.516489e-24
30	Средневзвешенный импакт-фактор журналов, в которых были процитированы статьи	0.778470	9.665581e-22

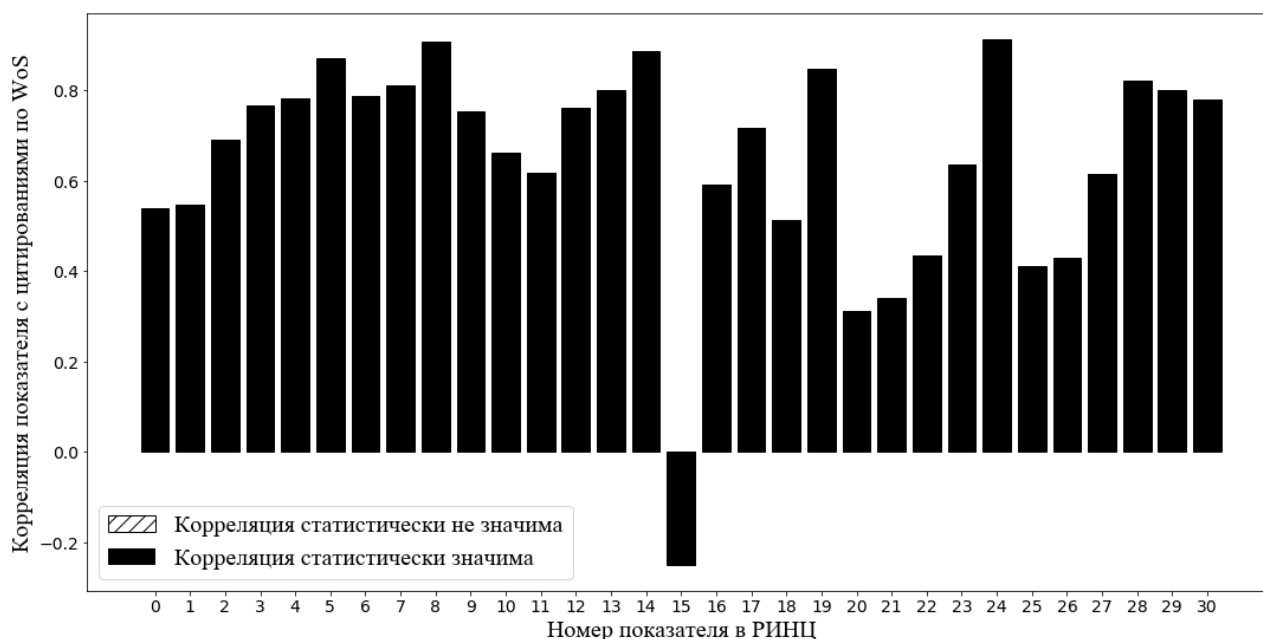


Рис. 1. Показатели в РИНЦ всех анализируемых исследователей

Как можно проинтерпретировать эти данные?

1. На первом месте «тавтологический» параметр №24 «Число цитирований из зарубежных журналов», что ожидаемо.
2. На втором месте корреляция (0,907) у параметра №8 «Индекс Хирша по ядру РИНЦ». Известно, что индекс Хирша часто оказывается хорошим предиктором, если база, по которому он считается довольно «сильная». Следовательно, база «ядро РИНЦ» весьма качественная. Для сравнения, индексы Хирша по РИНЦ и по elibrary.ru коррелируют с числом ссылок по WoS значительно хуже. Но, с другой стороны, индексы Хирша даже по таким слабым базам оказываются лучше многих других показателей. Это говорит о том, что индекс Хирша даже на относительно слабой базе может получить приемлемый результат.
3. На третьем месте (0,888) показатель №14 «Индекс Хирша с учетом только статей в журналах». Это говорит о том, что если мы убираем из рассмотрения монографии и оставляем только лишь статьи, то это позволяет сконструировать хороший показатель даже на относительно слабой базе.
4. На четвертом месте (0,870) показатель №5 «Число цитирований из публикаций, входящих в ядро РИНЦ». То, что цитирования по ядру РИНЦ хорошо коррелируют с цитированиями из WoS даже без применения индекса Хирша, говорит о том, что ядро РИНЦ — неплохая база.
5. Параметры, отражающие число статей по различным базам, коррелируют с числом ссылок по WoS следующим образом: elibrary.ru (0,538), РИНЦ (0,548), ядро РИНЦ (0,690). Это аргумент в пользу того, что силы баз расположены именно в таком порядке. Аналогично, параметры, отражающие число ссылок из статей по этим же базам, ведут себя также — elibrary.ru (0,767); РИНЦ (0,782); ядро РИНЦ (0,871). Также аналогично индексы Хирша по elibrary.ru (0,786), по РИНЦ (0,811), по ядру РИНЦ (0,907). То, что разным показателям — число статей, число ссылок и индексы Хирша — значения корреляций по разным базам упорядочено одинаково, это довольно сильные аргументы в пользу того, что РИНЦ действительно сильнее elibrary.ru, а

ядро РИНЦ сильнее РИНЦа. Это, в частности, говорит о том, что процедура исключения недобросовестных журналов из базы РИНЦ работает довольно неплохо.

6. Показатели исключительно «интенсивные», завязанные на импакт-фактор №29 «Средневзвешенный импакт-фактор журналов, в которых были опубликованы статьи» и №30 «Средневзвешенный импакт-фактор журналов, в которых были процитированы статьи» находятся на 9-м и 12-м местах соответственно. Это говорит о том, что в целом, когда в группе есть и очень плодовые исследователи-корифеи, «интенсивные» показатели не являются очень сильными предикторами.
7. Единственная отрицательная корреляция у параметра №15 «Год первой публикации» (-0.250). Это выглядит вполне естественно: чем позже исследователь начал писать научные статьи, тем меньше у него цитирований из WoS.

3 Анализ исследователей «среднего эшелона»

Существует ли такое подмножество анализируемого множества исследователей, для которого какие-либо показатели, кроме показателя №15, имеют статистически значимую отрицательную корреляцию с числом ссылок из WoS? Почему это может быть важно? Например, в работе [4] показано, что привнесение в расчеты дополнительной информации, в которой эксперт не уверен, приводит к ошибкам отбора наилучшего варианта при многокритериальном выборе. На более высоком уровне абстракции можно сказать, что использование нерафинированных в том или ином смысле исходных данных приводит к вполне измеримым ухудшениям экстремизируемого показателя. Поэтому и здесь естественно предположить, что некоторая гетерогенность датасета, если её не учитывать, может приводить к снижению точности моделей, обученных на нем. Под гетерогенностью датасета в данном случае понимаем ситуации, когда показатель с целевой переменной положительно скоррелирован на всем датасете и отрицательно на некотором его подмножестве. Кроме того, существуют ли подмножества, для которых наибольшее значение корреляции имеют «интенсивные» показатели?

В общем случае ответ на эти вопросы найти вычислительно сложно. Число всех подмножеств данного множества равно 2^N , где N — число элементов в множестве, так что время поиска подмножеств, удовлетворяющих необходимым условиям, при переборе «грубой силой» будет расти экспоненциально с ростом числа элементов. Можно ли каким-то образом снизить вычислительную сложность поиска подмножеств с заданными свойствами — это предмет отдельного исследования. Поэтому мы пока воспользуемся эвристическим соображением и проанализируем отдельно подмножества исследователей, имеющих среднее и большое число статей в ядре РИНЦ.

Итак, рассмотрим «крепких середнячков» — исследователей, имеющих некое среднее количество публикаций в ядре РИНЦ, допустим, от 30 до 100 включительно. Из анализируемых 101 исследователей, имеющих цитирования по базе WoS, указанное число работ в ядре РИНЦ имеют 28 исследователей. Как и для выборки выше, по данной выборке для каждого из параметров посчитаем корреляцию Спирмена с числом цитирований по WoS (табл. 2, рис. 2).

Таблица 2. Показатели в РИНЦ исследователей, имеющих от 30 до 100 работ в ядре РИНЦ

№	Наименование показателя в РИНЦ	Корреляция	р-критерий
0	Число публикаций на elibrary.ru	-0.371988	0.051268
1	Число публикаций в РИНЦ	-0.361134	0.059012
2	Число публикаций, входящих в ядро РИНЦ	0.581826	0.001163
3	Число цитирований из публикаций на elibrary.ru	0.200903	0.305311
4	Число цитирований из публикаций, входящих в РИНЦ	0.224716	0.250288
5	Число цитирований из публикаций, входящих в ядро РИНЦ	0.656036	0.000150
6	Индекс Хирша по всем публикациям на elibrary.ru	0.220651	0.259185
7	Индекс Хирша по публикациям в РИНЦ	0.235499	0.227669
8	Индекс Хирша по ядру РИНЦ	0.644292	0.000215
9	Число публикаций, процитировавших работы автора	0.126762	0.520365
10	Число ссылок на самую цитируемую публикацию	0.222831	0.254389
11	Число публикаций автора, процитированных хотя бы один раз	-0.119726	0.543960
12	Среднее число цитирований в расчете на одну публикацию	0.491583	0.007890
13	Индекс Хирша без учета самоцитирований	0.271672	0.161970
14	Индекс Хирша с учетом только статей в журналах	0.555472	0.002151
15	Год первой публикации	0.049499	0.802484
16	Число самоцитирований	0.153551	0.435322
17	Число цитирований соавторами	0.267415	0.168909

№	Наименование показателя в РИНЦ	Корреляция	p-критерий
18	Число соавторов	-0.018072	0.927273
19	Число статей в зарубежных журналах	0.595972	0.000818
20	Число статей в российских журналах	-0.526085	0.004034
21	Число статей в российских журналах из перечня ВАК	-0.432307	0.021591
22	Число статей в российских переводных журналах	-0.060830	0.758468
23	Число статей в журналах с ненулевым импакт-фактором	-0.043031	0.827881
24	Число цитирований из зарубежных журналов	0.755714	0.000003
25	Число цитирований из российских журналов	-0.471804	0.011254
26	Число цитирований из российских журналов из перечня ВАК	-0.395237	0.037373
27	Число цитирований из российских переводных журналов	0.212673	0.277240
28	Число цитирований из журналов с ненулевым импакт-фактором	0.429725	0.022475
29	Средневзвешенный импакт-фактор журналов, в которых были опубликованы статьи	0.680717	0.000067
30	Средневзвешенный импакт-фактор журналов, в которых были процитированы статьи	0.669495	0.000098

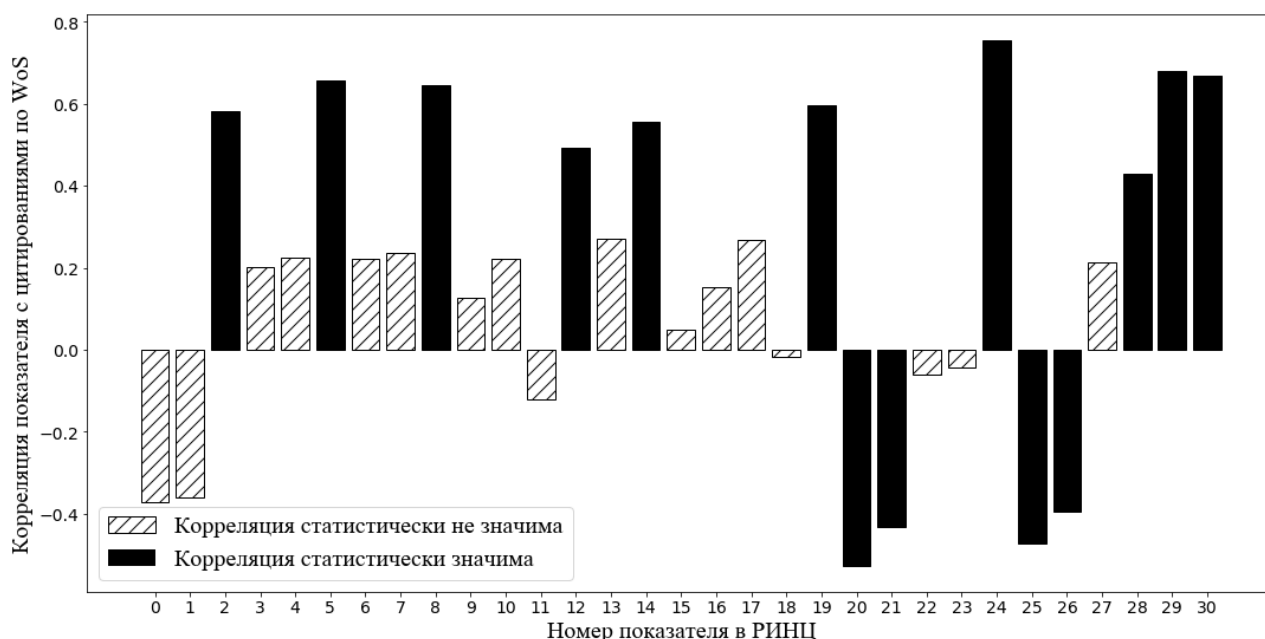


Рис. 2. Показатели в РИНЦ исследователей, имеющих от 30 до 100 работ в ядре РИНЦ
Как можно проинтерпретировать эти данные?

1. На первом месте «тавтологический» параметр №24 «Число цитирований из зарубежных журналов», что ожидаемо.
2. А вот на втором месте «интенсивный» показатель №29 «Средневзвешенный импакт-фактор журналов, в которых были опубликованы статьи». Получается, начинающим исследователям, стремящимся перейти в средний эшелон, имеет смысл ориентироваться на качество журналов, где они собираются публиковаться, отраженное в импакт-факторе. А не на количество статей.
3. На третьем месте показатель №30 «Средневзвешенный импакт-фактор журналов, в которых были процитированы статьи». Интерпретация здесь та же: если ориентироваться на качественные журналы, то и ссылки будут, скорее всего, из качественных журналов.
4. На четвертом месте показатель №5 «Число цитирований из публикаций, входящих в ядро РИНЦ». Здесь стоит отметить два момента. Во-первых, показатель основывается на ядре РИНЦ, что в очередной раз говорит, ядро РИНЦ — неплохая база. А во-вторых, на этой выборке параметр №5 опережает параметр №8 «Индекс Хирша по ядру РИНЦ». Это позволяет предположить, что для не суперплодовитых авторов, возможно, стоит ориентироваться на число ссылок из сильной базы. А вот для корифеев (и это будет некоторым образом подтверждено ниже) имеет смысл ориентироваться на индекс Хирша по сильной базе.
5. Мы видим, что есть четыре параметра — №№20, 21, 25 и 26 — по которым наблюдается небольшая, но тем не менее существенная статистически значимая отрицательная корреляция. Все они связаны с российскими и, в частности, ВАКовскими журналами. Наибольшей по

- абсолютному значению отрицательной корреляцией обладает показатель №20 «Число статей в российских журналах» (-0.526). Получается, что избыточная гонка за числом статей в журналах, не являющихся международными, идет вразрез с получением цитирований по WoS.
6. На втором месте с конца показатель №25 «Число цитирований из российских журналов» (-0.472). Видим, что, при прочих равных, цитирования — часто менее манипулируемый параметр, чем число статей.
 7. На третьем месте с конца показатель №21 «Число цитирований из российских журналов из перечня ВАК» (-0.432). Это говорит о том, что хоть и некоторые усилия по формированию перечня ВАК видны (всё-таки отрицательная корреляция меньше по абсолютному значению, чем у просто российских журналов), но тем не менее эти усилия оказываются недостаточными, чтобы база ВАК могла считаться хорошей: как и в случае с российскими журналами, ориентация на ВАКовский список скорее уводит в сторону от международного признания, чем помогает продвинуться вперед.
 8. На четвертом месте с конца показатель №26 «Число цитирований из российских журналов из перечня ВАК» (-0.395). Здесь можно сделать те же выводы, что и раньше. Во-первых, цитирования при прочих равных оказывается менее манипулируемым параметром, чем число статей. И во-вторых, ВАКовский список несколько чище, чем перечень всех российских журналов, но всё же не настолько, чтобы на него можно было ориентироваться. Более того, ориентироваться на него просто вредно.
 9. Видим, что параметр №15 «Год первой публикации» статистически значимо не коррелирует с числом ссылок по WoS. Это говорит о том, что для этой группы еще не очень возрастных исследователей среднего эшелона не так важно, чуть раньше или чуть позже автор начал писать. Важно, в журналах с каким импакт-фактором он публикуется.
 10. Корреляции показателей №0 «Число публикаций на elibrary.ru» и №1 «Число публикаций в РИНЦ» с числом ссылок из WoS немного не проходят уровень статистической значимости, но, наверное, если бы данных было больше, то оказалось бы статистически значимо, что ориентироваться на число публикаций как в РИНЦ, так и публикаций на сайте elibrary.ru вообще не стоит: такая ориентация уведет в противоположном направлении от международного признания.

4 Анализ корифеев

Теперь проанализируем исследователей, имеющих более 200 работ в ядре РИНЦ. Таких исследователей в анализируемом датасете 39.

Таблица 3. Показатели в РИНЦ исследователей, имеющих более 200 работ в ядре РИНЦ

№	Наименование показателя в РИНЦ	Корреляция	p-критерий
0	Число публикаций на elibrary.ru	0.300552	6.301006e-02
1	Число публикаций в РИНЦ	0.304788	5.920913e-02
2	Число публикаций, входящих в ядро РИНЦ	0.431297	6.119564e-03
3	Число цитирований из публикаций на elibrary.ru	0.784818	3.348428e-09
4	Число цитирований из публикаций, входящих в РИНЦ	0.802632	7.967975e-10
5	Число цитирований из публикаций, входящих в ядро РИНЦ	0.841093	2.044595e-11
6	Индекс Хирша по всем публикациям на elibrary.ru	0.837151	3.105763e-11
7	Индекс Хирша по публикациям в РИНЦ	0.843855	1.515144e-11
8	Индекс Хирша по ядру РИНЦ	0.806555	5.695483e-10
9	Число публикаций, процитировавших работы автора	0.720648	2.298869e-07
10	Число ссылок на самую цитируемую публикацию	0.519992	6.915802e-04
11	Число публикаций автора, процитированных хотя бы один раз	0.497115	1.284273e-03
12	Среднее число цитирований в расчете на одну публикацию	0.736032	9.322560e-08
13	Индекс Хирша без учета самоцитирований	0.778669	5.328819e-09
14	Индекс Хирша с учетом только статей в журналах	0.843013	1.661144e-11
15	Год первой публикации	-0.316323	4.977771e-02
16	Число самоцитирований	0.369636	2.056001e-02
17	Число цитирований соавторами	0.611741	3.507704e-05
18	Число соавторов	0.258757	1.117031e-01
19	Число статей в зарубежных журналах	0.690996	1.119295e-06
20	Число статей в российских журналах	-0.002227	9.892644e-01
21	Число статей в российских журналах из перечня ВАК	-0.002936	9.858496e-01

№	Наименование показателя в РИНЦ	Корреляция	p-критерий
22	Число статей в российских переводных журналах	0.018930	9.089352e-01
23	Число статей в журналах с ненулевым импакт-фактором	0.458626	3.320131e-03
24	Число цитирований из зарубежных журналов	0.847773	9.805010e-12
25	Число цитирований из российских журналов	0.165385	3.143321e-01
26	Число цитирований из российских журналов из перечня ВАК	0.155980	3.430173e-01
27	Число цитирований из российских переводных журналов	0.415789	8.475747e-03
28	Число цитирований из журналов с ненулевым импакт-фактором	0.821255	1.507942e-10
29	Средневзвешенный импакт-фактор журналов, в которых были опубликованы статьи	0.573077	1.372692e-04
30	Средневзвешенный импакт-фактор журналов, в которых были процитированы статьи	0.587854	8.317455e-05

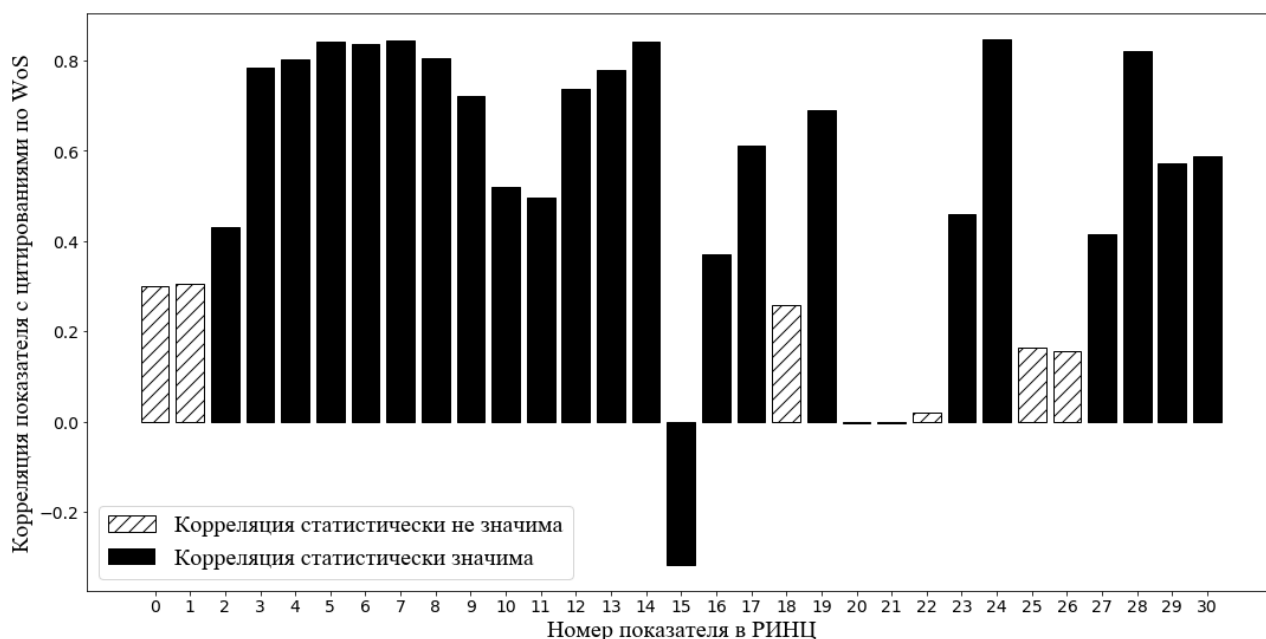


Рис. 3. Показатели в РИНЦ исследователей, имеющих более 200 работ в ядре РИНЦ

Как можно проинтерпретировать эти данные?

1. На первом месте «тавтологический» параметр №24 «Число цитирований из зарубежных журналов», что ожидаемо. Стоит иметь ввиду, что различия между нижеследующими показателями-лидерами невелико.
2. На втором месте параметр №7 «Индекс Хирша по публикациям в РИНЦ». Выглядит так, как будто если все анализируемые имеют сопоставимо большое число публикаций, то индекс Хирша является довольно сильным предиктором даже на относительно «слабой» и шумной базе РИНЦ.
3. На третьем месте параметр №14 «Индекс Хирша с учетом только статей в журналах». То есть, отделение статей от монографий увеличивает предсказательную способность показателя. По-видимому, это связано с тем, что есть «чрезмерно популярные» исследователи: на них принято ссылаться даже среди тех, кто их работы детально не изучал. В итоге исследователи не знают, на какой именно тезис «популярных корифеев», представленный в некоторой статье, стоит сослаться, поэтому ссылаются на одну из самых известных монографий корифеев.
4. На четвертом месте параметр №5 «Число цитирований из публикаций, входящих в ядро РИНЦ». Это ожидаемо, учитывая проведенный выше анализ других подвыборок исследователей.
5. «Интенсивные параметры» №29 «Средневзвешенный импакт-фактор журналов, в которых были опубликованы статьи» и №30 «Средневзвешенный импакт-фактор журналов, в которых были процитированы статьи» значительно отстают от показателей-лидеров по предсказательной способности. Это говорит о том, что чтобы из среднего эшелона исследователей перейти в корифеи придется не только ориентироваться на «плотность дискурса», но и обращать внимание на количество написанных статей.

6. Показатели №№ 20, 21, 25, 26, связанные с исключительно российскими журналами и журналами из списка ВАК, статистически значимо не коррелируют с числом ссылок из WoS.
7. Показатель №15 «Год первой публикации» имеет небольшую (-0.316), но статистически значимую отрицательную корреляцию. Интерпретация очевидна: чем позже начал писать, тем при прочих равных меньше ссылок по WoS.
8. Несмотря на то, что выборки исследователей среднего эшелона и «корифеев» не сильно отличаются по объему, у «корифеев» гораздо больше статистически значимо коррелирующих показателей (23 против 14). Наверное, это можно проинтерпретировать так, что корифеи похожи друг на друга существенно больше, чем «среднячки» друг на друга.

Заключение

Если не делить анализируемых исследователей на группы, то самыми сильными предикторами числа их цитирований по WoS в порядке убывания будут: число цитирований из зарубежных журналов; индекс Хирша по ядру РИНЦ; число цитирований из ядра РИНЦ. То есть, во-первых, «тавтологический» показатель «число ссылок из зарубежных журналов» является действительно сильным предиктором. Это говорит о том, что какой бы шумной базой РИНЦ ни была, ей всё же хватает чистоты для того, чтобы этот предиктор был сильным. А во-вторых, ядро РИНЦ достаточно сильное и чистое, чтобы начали действовать общеизвестные закономерности: чтобы были показательны индекс Хирша по этой базе (ядру РИНЦ) и число цитирований по этой базе. Если выделить группу «ученых-корифеев», то всё будет так же, только более выражено.

Если выделить группу ученых среднего эшелона, то самым сильным предиктором — после «тавтологического» показателя «число ссылок из зарубежных журналов» — будет средний импакт-фактор журналов, в которых опубликованы статьи. Это говорит о том, что для начинающих и «продолжающих» исследователей ориентация в своем развитии на «плотность дискурса» важнее, чем число статей, в том числе, и в ядре РИНЦ.

Число статей в российских и ВАКовских журналах, а также количество ссылок из них статистически значимо отрицательно коррелируют с числом цитирований по WoS. То есть выступают сильными отрицательными предикторами: при прочих равных для ученых среднего эшелона большее число публикаций в российских журналах говорит о том, что, вероятно, ссылок из WoS у него будет меньше.

То есть гонка за российскими публикациями идет вразрез с получением результатов, признаваемых международным сообществом, так как побуждает начинающих исследователей удаляться от достижений международной науки, от её методологии, от её принципов работы с материалом.

Есть работы, показывающие, что статьи из списка ВАК (без ядра РИНЦ) цитируются почти также низко, как и статьи, входящие в РИНЦ, но не входящие в ядро РИНЦ [5]. В данной работе показано, что ориентироваться на него не только не очень полезно, но иногда даже контрпродуктивно.

Литература

1. Салтыков С.А., Русяева Е.Ю. Медиация в науковедении: экспертно-наукометрический подход // Проблемы управления. 2017. №6. С. 63-67.
2. Салтыков С.А. Анализируемый датасет. URL: https://github.com/sergiustoday/rints-wos/blob/master/RintsWos1_1000.pkl (дата обращения: 29.04.2020).
3. Пресс-релиз Рабочей группы по оценке качества и отбору журналов в Russian Science Citation Index (RSCI) на платформе Web of Science. URL: https://elibrary.ru/projects/rsci/rsci_press2020.pdf (дата обращения: 29.04.2020).
4. Салтыков С.А. Экспериментальное сопоставление методов взвешенной суммы, теории полезности и теории важности критериев для решения многокритериальных задач с балльными критериями // Управление большими системами. 2010. № 29. С. 16-41.
5. Сравнение уровня публикаций российских ученых в базах данных Web of Science, Scopus и RSCI // Научная электронная библиотека. URL: https://www.elibrary.ru/wos_scopus_rsci.asp (дата обращения: 29.04.2020).