
КОНСУЛЬТАЦИИ

О.В. Терещенко
(Минск)

МЕТОД КОГОРТНОГО АНАЛИЗА В СОЦИАЛЬНЫХ ИССЛЕДОВАНИЯХ

Статья посвящена применению метода когортного анализа в социальных исследованиях. В ней в форме ответов на вопросы рассматривается роль метода в повторных мониторинговых и лонгитюдных исследованиях, его преимущества при изучении изменений в социальном поведении различных возрастных групп с течением времени.

Ключевые слова: когорта, когортный анализ, эффект когорты, эффект возраста, эффект времени, мониторинг, панельные исследования, лонгитюдные исследования.

Что такое когортный анализ? Когортный анализ – это сравнительный анализ изменений в социальном поведении одной или нескольких когорт с течением времени. Слово *когорта* латинского происхождения (*cohort*) и первоначально означало военный отряд, составлявший десятую часть римского легиона. Позднее его стали использовать для обозначения сплоченной группы единомышленников (в толковых словарях встречается пример «ленинская когорта»). В научный оборот термин «когорта» введен в 1947 г. П. Уэлптоном для обозначения совокупности людей, у которых в один и тот же период времени произошло определенное *демографическое событие* [1, с. 182]. Например, когортой

Ольга Викентовна Терещенко – кандидат социологических наук, зав. кафедрой социальной коммуникации факультета философии и социальных наук Белорусского государственного университета. E-mail: oteresch@tut.by.

является группа лиц, родившихся в 1961–1965 гг., или группа женщин, вступивших в первый брак в 1980 г. Интервал времени, выбираемый для выделения когорты (один год или несколько лет), зависит от целей анализа и особенностей исходных данных. Исторически первым видом когортного анализа был *«продольный»* демографический анализ, направленный на изучение частоты событий (демографических), происходящих в когорте в зависимости от ее «возраста», т. е. от промежутка времени между ее образованием и изучаемым событием – смертью, вступлением в брак, разводом, рождением первого ребенка и последующих детей и т.п.

В современных реалиях когортный анализ как метод используется не только в демографии, но и в других отраслях знания. Например, в социологии широкое распространение получили лонгитюдные исследования жизненного пути образовательных когорт, выделенных по критерию завершения заданного уровня образования в определенном году (см., например, [2]). В качестве когорты может рассматриваться группа предприятий, созданных в один и тот же год или ограниченный временной период (например, в годы перестройки), автомобили одной модификации, вина урожая одного года и т.п.

Методологические и функциональные возможности метода постоянно расширяются. Если *«продольный»* анализ изучал одну когорту и частоту происходящих в ней событий, то современный когортный анализ позволяет сравнивать жизненные пути нескольких когорт, а также возрастные и временные аспекты изучаемых процессов. В любом случае исследование с использованием этого метода является достаточно продолжительным (сопоставимым по длительности с жизнью изучаемых объектов), в значительной мере ретроспективным (может быть осуществлено не ранее, чем событие наступит хотя бы для части изучаемой когорты) и методологически сложным, так как предполагает разграничение эффектов возраста, когорты и времени.

Оправданы ли такие сложности? Разве возраст сам по себе (наряду с полом, образованием и др.) не объясняет различия в поведении людей, организаций, тех же автомобилей? В том-то и дело, что возраст – чрезвычайно коварная объясняющая переменная. Когда мы проводим сравнение по возрасту, возникает явление, получившее в экспериментальных исследованиях название *смешения эффектов разных факторов*. В случае когортного анализа – это эффекты трех факторов: возраста, когорты и времени. *Фактор возраста* соотносится с такими свойствами личности как физическое здоровье, степень мобильности, интерес ко всему новому и способность его восприятия и т.д. *Фактор когорты* связан, в первую очередь, с условиями ее возникновения, становления, социализации, жизненного опыта, доступа к различного рода ресурсам и т.п. Например, пожилые люди значительно хуже относятся к телевизионной рекламе, чем молодые, не только потому, что нелюбопытны и им не нравится агрессивный стиль рекламных роликов (эффект возраста), но также потому, что вся прожитая жизнь приучила их к экономии и осторожности в принятии решений о покупках (эффект когорты). *Фактор времени* отражает перемены, происходящие в макросреде – экономические, политические, социальные, информационные и др. Например, с помощью последовательной информационной политики, разъясняющей необходимость рекламы для развития экономики, можно добиться снижения неприятия рекламы во всех возрастных когортах.

Замечательный пример того, как опасно строить на данных единовременных исследований («срезов») серьезные выводы и прогнозы, приводит Р. Дэвис в статье «От срезового к лонгитюдному анализу» [3, р. 24]. Представим себе, что сервисный центр ведет статистику надежности (индикатор – число дней простоя в год) стоящих на обслуживании автомобилей с группировкой по возрастам (табл. 1).

Таблица 1

ЧИСЛО ДНЕЙ ПРОСТОЯ В ЗАВИСИМОСТИ ОТ ВОЗРАСТА
АВТОМОБИЛЯ

Индикатор	Возраст автомобиля, лет				
	1	2	3	4	5
Среднее число дней простоя в год	4	3	15	16	18

Наиболее простым объяснением вполне очевидной зависимости является эффект возраста: чем старше автомобиль, тем он менее надежен. Если это так, то можно предположить, что в последующие годы автомобили будут «вести себя» аналогично, бизнес можно планировать как устойчивый и рассчитывать на тот же уровень прибыли. Но специалисту, знакомому с эффектами когорты, столь же очевидным представляется и *альтернативное объяснение*: два года назад сменилось поколение автомобилей. Новые автомобили (которым в данный момент 1-2 года) будут и впредь надежными, автомобили старшего поколения (3 года и старше) изначально были ненадежными и таковыми останутся. Поскольку с каждым годом их будет становиться меньше, для поддержания бизнеса (не говоря уже о его развитии) понадобятся свежие решения.

Но как же понять, с каким из случаев мы имеем дело? По данным одного «среза» это сделать невозможно и необходим сравнительный анализ трендов для автомобилей каждого года выпуска, т.е. когортный анализ, как минимум, за последние пять лет.

Чтобы осуществить его, имеющуюся статистику следует представить в виде таблицы, строки которой будут соответствовать возрасту автомобилей, а столбцы – годам наблюдения. В клетке таблицы, образованной пересечением строки и столбца, указывается усредненное значение изучаемого показателя для соответствующей когорты в соответствующем году. Таковую таблицу читают по диагонали, по мере «старения» когорты и течения времени.

В нашем примере данные, соответствующие гипотезе об эффекте возраста, могли бы выглядеть, как показано в табл. 2. Ко-

горта автомобилей выпуска 1986 г. в первый год работы простаивала в среднем 4 дня, во второй (1987 г.) – 3 дня, на третьем году эксплуатации (1988 г.) надежность резко снизилась – 17 дней простоя и с тех пор остается на том же уровне: в 1989 г. – 15 дней, в 1990 г. – 18 дней. Заметим, что столбец, соответствующий 1990 г., полностью совпадает с распределением из табл. 1.

Таблица 2

ЧИСЛО ДНЕЙ ПРОСТОЯ В ЗАВИСИМОСТИ ОТ ВОЗРАСТА
АВТОМОБИЛЕЙ (ЭФФЕКТ ВОЗРАСТА)

Возраст автомобиля, лет	Год выпуска					Когорта (год образования)
	1986	1987	1988	1989	1990	
5					18	1986
4				15	16	1987
3			17	14	15	1988
2		3	4	4	3	1989
1	4	3	3	3	4	1990

Если же верна гипотеза об эффекте когорты, то взятые в ретроспективе данные могут выглядеть, например, следующим образом (табл. 3). Соответственно в последнем столбце показатели те же.

Таблица 3

ЧИСЛО ДНЕЙ ПРОСТОЯ В ЗАВИСИМОСТИ ОТ ВОЗРАСТА
АВТОМОБИЛЕЙ (ЭФФЕКТ КОГОРТЫ)

Возраст автомобиля, лет	Год выпуска					Когорта (год образования)
	1986	1987	1988	1989	1990	
5					18	1986
4				15	16	1987
3			16	17	15	1988
2		14	18	15	3	1989
1	17	15	16	4	4	1990

Из табл. 3 видно, что когорты автомобилей 1986-го, 1987-го и 1988 гг. выпуска с первого года эксплуатации показывали ту же надежность, что и в текущем 1990 г., а именно 14-17 дней простоя в год. Автомобили когорт 1989-го и 1990 гг. изначально имеют другую надежность, и нет никаких оснований предполагать, что в возрасте 3-5 лет они будут столь же часто нуждаться в ремонте, как автомобили предшествующего поколения. Впрочем пока нет данных, подтверждающих, что их надежность с возрастом не уменьшится. Поэтому исследование необходимо продолжить!

Получается, что когортный анализ – это просто форма представления статистических данных за несколько лет? Отчасти да. Когортный анализ действительно «пришел» из демографии, где возрастные когорты традиционно сравниваются, например, по продолжительности жизни и другим показателям, полученным с помощью текущего статистического учета и в ходе переписей населения. Рассмотренный выше пример из статьи Р. Дэвиса также базируется на внутренней статистике фирмы.

Однако применение метода когортного анализа в социологии и маркетинге было бы чрезвычайно ограничено, если бы он мог осуществляться только на полных данных о генеральной совокупности. Метод в простейшем его виде может быть реализован и на данных регулярных дескриптивных обследований по репрезентативной выборке, которая в ходе обработки данных разделяется на возрастные группы, и для каждой группы вычисляется среднее арифметическое изучаемого показателя. В этом случае условием корректности сравнений является неизменность генеральной совокупности и методов измерения «трендовых» показателей, а также схемы формирования выборки. Кроме того, для обнаруженных различий необходимо проверять их статистическую значимость.

Ретроспективный когортный анализ данных выборочно-повторного исследования был предпринят в начале 1980-х гг. Дж. Ренцом, Ф. Рейнолдсом и Р. Стаутом с целью опровергнуть бытовавший в то время среди маркетологов стереотип, соглас-

но которому употребление безалкогольных напитков с возрастом уменьшается [цит. по 4, с. 144]. Анализу были подвергнуты данные четырех маркетинговых опросов (1950-го, 1960-го, 1970-го и 1980 гг.), в каждом из которых респондентов спрашивали, употребляют ли они безалкогольные напитки. Выборка по каждому из опросов была сгруппирована по возрасту в 10-летние интервалы. В качестве трендового показателя для возрастной группы определялся процент респондентов, ответивших, что они употребляют безалкогольные напитки (табл. 4). По форме представления таблица напоминает план эксперимента с двумя независимыми переменными (факторами) – возраст и год исследования. На первый взгляд, гипотеза об эффекте возраста (снижении с возрастом потребления безалкогольных напитков) подтверждается в каждом из четырех опросов. Очевиден также эффект времени, проявляющийся в том, что от опроса к опросу процент потребителей безалкогольных напитков растет в каждой возрастной группе.

Таблица 4

РАСПРЕДЕЛЕНИЕ ОПРОШЕННЫХ ПО ПОТРЕБЛЕНИЮ
НАПИТКОВ В РАЗНЫЕ ГОДЫ,
% от доли в возрастной группе

Возраст, лет	Год опроса			
	1950	1960	1970	1980
18-19	52,9	62,6	73,2	81,0
20-29	45,2	60,7	76,0	75,8
30-39	33,9	46,6	67,7	71,4
40-49	23,2	40,8	58,6	67,8
50-59	18,1	28,8	50,0	51,9

При «продольном» когортном анализе читать ту же таблицу следует по диагонали, по мере течения времени и старения когорты. Каждая полная или неполная диагональ соответствует одной из когорт, выделяемых в процессе анализа. Табл. 5 получена из табл. 4 после выделения и обозначения когорт. Например, когор-

та № 4, данные о которой выделены в табл. 5 жирным шрифтом, в 1950 г. находилась в возрасте 20-29 лет и, соответственно, родилась в 1921–1930 гг. В 1950 г. 45,2% опрошенных представителей этой когорты ответили, что они употребляют безалкогольные напитки. В 1960 г., достигнув возраста 30-39 лет, в этой же когорте безалкогольные напитки употребляли 46,6% опрошенных, в 1970 г. в возрасте 40-49 лет – 58,6%, в 1980 г. в возрасте 50-59 лет – 51,9%.

Таблица 5

ПОТРЕБЛЕНИЕ БЕЗАЛКОГОЛЬНЫХ НАПИТКОВ:
«ПРОДОЛЬНЫЙ» КОГОРТНЫЙ АНАЛИЗ

Возраст, лет	Год опроса				Номер когорты
	1950	1960	1970	1980	
18-19	52,9	62,6	73,2	81,0	
20-29	45,2	60,7	76,0	75,8	8 (1961–1970)
30-39	33,9	46,6	67,7	71,4	7 (1951–1960)
40-49	23,2	40,8	58,6	67,8	6 (1941–1950)
50-59	18,1	28,8	50,0	51,9	5 (1931–1940)
Номер когорты		1 (1891–1900)	2 (1901–1910)	3 (1911–1920)	4 (1921–1930)

Разделение выборки каждого из четырех опросов на 10-летние возрастные группы позволяет выделить в этой таблице 8 возрастных когорт, наблюдение за которыми имело разную продолжительность. Так, когорта № 1 (1891–1900 гг. рождения) прослеживалась только в исследовании 1950 г. в возрасте 50-59 лет. Ее, конечно, нельзя подвергнуть «продольному» анализу, однако она обладает безусловной ценностью для сравнений с последующими когортами по мере достижения ими возраста 50-59 лет, а так-

же с другими возрастными в «срезовом» исследовании 1950 г. Аналогичным образом дело обстоит и с когортой №8 (1961–1970 гг. рождения), представители которой впервые появляются в опросе 1980 г. Остальные 6 представленных в таблице когорт позволяют проследить изменение потребительского поведения в отношении безалкогольных напитков на протяжении 10-30 лет (в зависимости от числа последовательных опросов, в которых они принимали участие). Четыре когорты (под номерами 2, 3, 5 и 7) убедительно показывают, что сформированная в том или ином возрасте привычка пить безалкогольные напитки с течением времени не исчезает, а число потребителей таких напитков в когорте растет. Исключение составляют когорты № 4, в которой снижается потребление безалкогольных напитков при переходе из возрастной группы «сорокалетних» (58,6%) в группу «пятидесятилетних» (51,9%), а также когорты № 6, в которой аналогичная тенденция наблюдается при переходе из группы «двадцатилетних» (76%) в группу «тридцатилетних» (71,4%). Узнать, являются эти отклонения следствием ошибки выборки или они говорят о некой новой тенденции, можно только продолжив исследования с прежней или даже более короткой периодичностью.

Тем самым, читая таблицу когортного анализа *по диагонали*, мы получаем данные о поведении когорты на протяжении определенного временного периода. Анализ таблицы *по столбцам* позволяет провести сравнение потребительского поведения различных возрастных групп в конкретный момент времени, например, с целью сегментирования рынка потребителей. *Строки таблицы* дают возможность проследить, как менялось с годами поведение каждой возрастной группы. Таким образом, когортный анализ позволяет разделить эффект когорты (по диагонали), эффект возраста (по столбцам) и эффект времени или социокультурных изменений (по строкам).

Когортный анализ эффективен на данных многолетних повторных исследований. Сколько замеров должно быть как ми-

нимум? Это зависит от объекта исследования, а также от того, насколько быстро протекают изучаемые процессы. Нередко даже относительно непродолжительные, но систематические наблюдения позволяют эффективно использовать метод. Проиллюстрируем это на примере когортного анализа белорусской интернет-аудитории по данным Независимого института социально-политических и экономических исследований [5] (см. табл. 6).

Таблица 6

РАСПРЕДЕЛЕНИЕ ИНТЕРНЕТ-АУДИТОРИИ ПО ВОЗРАСТУ,
% от доли пользователей в возрастной группе

Возраст, лет	Год		
	1998	2003	2008
16-19	3,3	44,1	71,0
20-24	4,5	43,4	75,7
25-29	8,0	33,8	61,3
30-34	2,4	23,3	48,5
35-39	2,4	17,7	48,6
40-44	1,1	17,7	43,2
45-49	1,1	11,9	36,8
50-54	0,5	9,8	30,3
55-59	0,5	2,9	10,5
60 и старше	0,3	2,8	3,6
Доля по выборке	2,3	17,8	31,2
Объем выборки	1474	1488	1531

Когортный анализ, в частности, показал, что первые Интернет-пользователи (в заметном количестве) появились в Беларуси в 1998 г. среди наиболее социально и экономически активной группы населения 25-29-летнего возраста. В настоящее время в этой когорте, достигшей 35-40-летнего возраста, пользователем Интернета является каждый второй. Она сохраняет преимущество перед старшими поколениями, что можно идентифицировать как эффект возраста. В то же время эта группа уступает более молодым, что,

безусловно, является эффектом когорты, так как они изначально находились в более выгодных условиях, во-первых, в силу многократного повышения экономической и технической доступности интернета, во-вторых, благодаря образовательной политике государства, обеспечившего Интернетом учебные заведения.

Основным недостатком представленных результатов является отсутствие статистической значимости различий между соседними возрастными группами внутри каждого временного периода. Это вызвано, прежде всего, использованием 5-летних возрастных интервалов, что обусловлено не только краткостью периода наблюдения, но и стремительностью развития процесса «интернетизации». С другой стороны, нами использованы данные, которые собирались не с целью последующего когортного анализа, а в ходе опросов общественного мнения. Выборка объемом 1500 респондентов вполне репрезентативно представляет десятиллионное население Беларуси в целом, но не достаточна для сравнительного анализа десяти 5-летних возрастных групп.

Как соотносится метод когортного анализа с лонгитюдными исследованиями? Лонгитюд – это панельное исследование, в котором выборка (*панель*) сформирована по критерию возраста. Этот подход предназначен для изучения жизненных путей возрастной когорты. Он ценится за уникальную возможность изучения изменений поведения респондентов, их ценностей и установок под влиянием тех или иных внешних событий [2]. Основным недостатком лонгитюдных исследований является очень высокая трудоемкость, связанная с необходимостью поддерживать контакты с респондентами на протяжении многих лет и обеспечивать их участие в очередных этапах, и, соответственно, высокая стоимость. Кроме того, систематические ошибки выборки, допущенные на первом этапе, даже если впоследствии они были исправлены, оказывают влияние на все последующие этапы лонгитюдного исследования. На репрезентативность поздних этапов лонгитюда влияет эффект вымирания панели, т.е. потеря значитель-

ного числа респондентов по причинам миграции, смертности, отказов от дальнейшего участия и т.д.

Сравнительный «продольный» анализ нескольких когорт может осуществляться как на основе данных параллельных лонгитюдов, идущих со «сдвигом» в несколько лет, так и на данных мониторинга – серии повторных исследований с независимыми выборками. Мониторинг значительно дешевле и организационно проще, чем лонгитюд, и построенные на его основе тренды менее подвержены систематическим ошибкам, однако он не обладает столь мощным аналитическим потенциалом. В частности, преимущество панельных исследований по сравнению с мониторинговыми Н. Малхотра демонстрирует следующим примером из маркетинговых исследований [4, с. 147]. В табл. 7 представлены результаты двух последовательных опросов, на основании которых можно заключить, что в промежуток между опросами ничего не изменилось, все торговые марки сохранили своих потребителей, следовательно, они удовлетворены и лояльны.

Таблица 7

РАСПРЕДЕЛЕНИЕ ПОКУПАЕМЫХ ТОРГОВЫХ МАРОК
(АБСОЛЮТНАЯ ЧАСТОТА)

Марка товаров	Первый опрос	Второй опрос
А	200	200
Б	300	300
В	500	500
Всего	1000	1000

Аналитические возможности мониторинга на этом исчерпаны. Но, если исследование было панельным, его данные могут быть дополнительно подвергнуты перекрестной табуляции (см. табл. 8). В этом случае становится очевидным, что численность потребителей не свидетельствует о сохранении их состава и, соответственно, удовлетворенности потребляемой маркой. По этой таблице можно

проанализировать не только распределение потребительских предпочтений, но и отношение к товарам конкурентов. При этом становится очевидным перемещение потребителей и тех, кто недоволен товаром (отток), и тех, кто недоволен товаром конкурентов (приток).

Таблица 8

ТАБЛИЦА СОПРЯЖЕННОСТИ ДВУХ ПАНЕЛЬНЫХ ОПРОСОВ

Марка товаров (первый опрос)	Марка товаров (второй опрос)			Всего
	А	Б	В	
А	100	50	50	200
Б	25	100	175	300
В	75	150	275	500
Всего	200	300	500	1000

Кроме того, специфика панели (отсутствие анонимности и наличие контактной информации) предоставляет дополнительную возможность выбрать респондентов, сменивших марку, и провести с ними глубинные интервью или фокус-группы для уточнения причин изменения потребительских предпочтений.

Что бы Вы посоветовали начинающим аналитикам? Думать о будущем. Выделять в текущих исследованиях вопросы, перспективные для образования трендов. Тщательно тестировать их формулировки. Повторять их время от времени, при удобном случае, по возможности, на выборках большого объема. Стараться сохранять формулировки, а при крайней необходимости изменений учитывать их влияние на долгосрочные тренды. Наконец, тривиальный совет: не группировать возраст в анкете, ибо это ограничивает возможность последующего выделения когорт.

ЛИТЕРАТУРА

1. Народонаселение: Энциклопедический словарь / Гл. ред. Г.Г. Меликьян. М.: Большая Российская энциклопедия, 1994.
2. Социальное расслоение возрастной когорты: Выпускники 80-х в постсоветском пространстве / Отв. ред. М.Х. Титма. М.: ИС РАН, 1997.

3. *Davis R.B.* From Cross-Sectional to Longitudinal Analysis // *Analyzing Social & Political Change: A Casebook of Methods* / Ed. by A. Dale, R.B. Davis. London: Sage Publications, 1994.

4. *Малхотра Н.* Маркетинговые исследования: практическое руководство. М.: Вильямс, 2007.

5. *Терещенко О.В.* Когортный подход к анализу белорусской аудитории интернета // Тезисы III Всероссийского социологического конгресса (Москва, 21-24 октября 2008 г.). М.: ИС РАН; РОС, 2008. (Диск CD, ISBN 978-5-89697-158-0).