

DOI: 10.14515/monitoring.2014.6.04

УДК 303.625:004.738.5

С.П. Моисеев, Ю.К. Савинкова

ВЫБОРКА, НАПРАВЛЯЕМАЯ РЕСПОНДЕНТОМ, В ОНЛАЙН-ОПРОСЕ: К ВОПРОСУ О ДИНАМИКЕ И КАЧЕСТВЕ

ВЫБОРКА, НАПРАВЛЯЕМАЯ
РЕСПОНДЕНТОМ, В ОНЛАЙН-ОПРОСЕ:
К ВОПРОСУ О ДИНАМИКЕ И КАЧЕСТВЕ

RESPONDENT-DRIVEN SAMPLING IN ONLINE
SURVEY: DYNAMICS AND QUALITY

МОИСЕЕВ Станислав Павлович — магистр социологии, менеджер исследовательских проектов в Институте фонда «Общественное мнение». E-mail: spmoiseev@gmail.com
САВИНКОВА Юлия Константиновна — магистр социологии, менеджер исследовательских проектов в компании Ipsos. E-mail: julia.savinkova@gmail.com

MOISEEV Stanislav Pavlovich – Master`s degree in Sociology, research project manager, Public Opinion Foundation Institute. E-mail: spmoiseev@gmail.com
SAVINKOVA Yuliya Konstantinovna - Master`s degree in Sociology, research project manager, Ipsos Institute. E-mail: julia.savinkova@gmail.com

Аннотация. Обсуждается¹ проблематика отбора респондентов для участия в онлайн-исследованиях. Проанализированы результаты и опыт использования выборки, направляемой респондентом (ВНР), для построения выборки в онлайн-опросах студентов НИУ ВШЭ. Проверено, соответствует ли онлайн-ВНР необходимым базовым допущениям: реципрокность связей между респондентами, случайный отбор и способность респондентов правильно оценить размер своей эго-сети. Дополнительный исследовательский вопрос касался изучения влияния знакомства «семян» (т.е. первых респондентов онлайн-ВНР) на динамику и качество итоговой выборки.

Для ответа на поставленные вопросы проведены два онлайн-опроса, основанные на использовании ВНР (онлайн-ВНР-1 и онлайн-ВНР-2). Ключевое отличие выборок состояло в том, что во втором случае «семена» отбирались из числа личных знакомых исследователей. На основе анализа результатов сделан вывод о том, что в случае

Abstract. The authors discuss the problems of selection of respondents for participation in online surveys. The results and the experience of the use of the respondent-driven sampling (RDS) used in online surveys among students of the Higher School of Economics were analyzed in the article. The authors checked whether the web-based RDS corresponds to the basic assumptions such as reciprocity of communications between respondents, random selection and respondent`s capability to adequately assess the size of the ego network. An additional question referred to the influence of the “seeds” (first RDS respondents) on the dynamics and the quality of the final sampling was also studied.

To answer the questions, two online RDS-based surveys were carried out by the author (online RDS-1 and online RDS-2). The key difference between these two studies is that the “seeds” in the second one were selected out of the researchers` own personal contacts. The authors conclude that the dynamics of the online RDS is more effective

¹ Используются результаты, полученные в ходе выполнения проекта No 13-05-0035, осуществляемого при поддержке Программы «Научный фонд НИУ ВШЭ» в 2013–2014 гг.

личного знакомства «семян» с исследователями динамика онлайн-ВНР оказывается эффективнее. При этом онлайн-ВНР-1 и онлайн-ВНР-2 не дали точной оценки контрольных параметров генеральной совокупности. Опираясь на полученные результаты, авторы формулируют задачи для дальнейших исследований в этом направлении.

Ключевые слова: RDS, WebRDS, онлайн-ВНР, методология онлайн-исследований, онлайн-опрос, сетевая выборка

if the “seeds” know the researcher in person. At the same time, online RDS-1 and online RDS-2 did not provide the exact assessment of the control parameters of the population. Based on the results of the study the authors define problems for future research.

Keywords: RDS, WebRDS, online survey methodology, web survey, network sample

В последние 15 лет в социальных науках растет интерес к использованию Интернета как среды и инструмента для проведения исследований, в частности онлайн-опросов. Эта практика открывает новые возможности и приводит к появлению новых проблем. К последним относятся сложности, связанные с вероятностным отбором респондентов, определением границ генеральной совокупности и оценкой статистической достоверности получаемых результатов [1, с. 17–18]. Перспективные способы их преодоления — использование выборки, направляемой респондентом (далее мы будем использовать аббревиатуру ВНР, которая соответствует обозначению RDS (Respondent-Driven Sample) в зарубежной литературе). ВНР — разновидность сетевой выборки, в которой исследователь определяет только первичную группу респондентов (так называемые «семена»), в то время как остальной объем выборочной совокупности формируется самими участниками опроса по заданным правилам. Изначально ВНР разрабатывалась для исследования закрытых и труднодоступных групп, однако несколько лет назад исследователи предложили использовать аналогичную технологию для отбора респондентов онлайн-опроса и назвали ее WebRSD (или онлайн-ВНР) [8]. Онлайн-ВНР основывается на двух теоретических моделях (цепи Маркова и модель гомофилии), позволяющих нивелировать негативные эффекты выборки снежного кома [5, 7]. Получаемая выборка не является случайной, однако она позволяет проводить статистическую оценку параметров генеральной совокупности [5–8].

Реализация ВНР опирается на 5 базовых допущений о генеральной совокупности, респондентах и процессе отбора [см., например: 4, 8]:

- 1 Респонденты поддерживают реципрокные (взаимные) отношения с людьми, которых они считают членами целевой группы.
- 2 Все респонденты принадлежат к одному компоненту сети.
- 3 Респондент может быть отобран для участия в исследовании несколько раз (повторно).
- 4 Респонденты могут точно оценить размер (степень) своей эго-сети.
- 5 В процессе рекрутинга респонденты совершают случайный выбор.

Предполагается, что при условии соблюдения всех правил и обозначенных выше допущений полученная сетевая выборка, реализованная в традиционном или онлайн-опросе, оказывается несмещенной [4, 7]. Эта гипотеза несколько раз проверялась и критически оценивалась в ходе экспериментов [3], однако исследователи по-прежнему не

пришли к согласованному выводу о том, как и в каких условиях ВНР позволяет получить заявленный статистически проверяемый и репрезентативный результат. Оппоненты Хекаторна заявляют о том, что сформулированные им и его коллегами допущения нереалистичны и ставят под сомнения качество всех исследований, в основании которых лежит ВНР [3]. Эта критика находит подтверждение не только в математических моделях, но и в эмпирических экспериментах. В частности исследование А.М. Мавлетовой показало, что при поведении онлайн-ВНР «не было подтверждено базовое допущение... о случайности рекрутирования участников» [2, с. 47]. Еще одна сложность, которая возникает при реализации ВНР, связана с отношениями между исследователем и «семенами»: на данный момент неизвестно, как на результаты ВНР влияет знакомство «семян» с исследователем и их лояльность к проекту и его автору.

В связи с этим мы ставим перед собой два исследовательских вопроса, адресованных к возможности применения ВНР в онлайн-исследованиях:

- 1 Можно ли построить онлайн-ВНР выборку студентов университета, которая бы соответствовала базовым предположениям ВНР?
- 2 Зависит ли успех проведения ВНР опроса от личного знакомства между «семенами» и исследователем?

Основная гипотеза состоит в том, что эмпирически проверяемые допущения об ВНР (1, 4, 5) окажутся соблюдены², а личное знакомство «семян» с исследователем будет иметь положительное влияние на скорость построения выборки и близость оцениваемых параметров выборочной совокупности к распределению этих же признаков в генеральной.

Объект исследования — студенты Национального исследовательского университета «Высшая школа экономики», которые стали участниками двух онлайн-опросов с реализацией выборки, направляемой респондентом.

Дизайн исследования

В основу дизайна проведенного исследования лег опыт Хекаторна и его соавторов, проводивших два онлайн-опроса среди студентов университета Корнелла, основанные на ВНР в 2004 и 2008 гг. [8, 9]. В обоих случаях опрос начинался с отбора 9 «семян», которые должны были заполнить онлайн-анкету и пригласить к участию в исследовании до 3-х студентов своего университета. Все коммуникации с респондентами осуществлялись по электронной почте. За заполнение анкеты и приглашение каждого из 3-х участников все респонденты получали финансовое вознаграждение. Итоговая выборка достигла 159 респондентов в 2004 г. и 369 респондентов в 2008 г.

С опорой на эту логику в данном исследовании с февраля по июнь 2013 г. предпринята попытка последовательного построения двух онлайн-ВНР (табл. 1). Для первой

² Одним из недостатков модели ВНР Д. Хекаторна и его коллег является то, что некоторые допущения, лежащие в ее основе, не подлежат прямой эмпирической проверке. В частности это касается предположения о том, что все респонденты принадлежат к одному компоненту сети (2) и могут быть отобраны для участия в исследовании повторно (3) [4, 8]. Положение № 2 обычно обосновывается с позиции теорий «Маленького мира», в то время как предположение № 3 можно считать соблюденным, если исследователь не будет исключать из анализа данных кейсы с повторным приглашением участников. В данной работе мы не ставим цели оспаривать эти суждения и принимаем их как аксиоматические.

выборки набраны респонденты, являющиеся так называемыми «звездами» среди студентов НИУ ВШЭ (они имели большое число социальных связей). Во вторую группу набраны «семена», лично знакомые с исследователями. Все респонденты, прошедшие опрос и пригласившие 3-х друзей, получали финансовое вознаграждение в размере 150 руб. Для построения сетевой выборки участников просили оставить данные – телефон, e-mail, имя, фамилию и данные человека, от которого было получено приглашение.

Таблица 1 Параметры выборок двух опросов

Параметр	Выборка опроса 1	Выборка опроса 2
Количество респондентов	72	79
Количество семян	6	5
Структура семян по полу м/ж	3/3	3/2
Наличие вознаграждения	Да	Да
Мотивация семян	Заинтересованность в исследовании/вознаграждении	Личная мотивация/ответственность перед исследователем

Для контроля параметров выборок использованы данные Центра внутреннего мониторинга НИУ ВШЭ (далее – ЦВМ) о распределении генеральной совокупности студентов по полу, факультетам, условиям проживания и занятости. Для сбора данных онлайн-опросов использовалось программное обеспечение «Unipark» (<http://www.unipark.com>).

Онлайн-ВНР-1

Для набора наиболее активных студентов на главном сайте НИУ ВШЭ и в группах НИУ ВШЭ в социальной сети «ВКонтакте» было размещено объявление с информацией об условиях участия в исследовании и вознаграждении. На него откликнулись 20 студентов НИУ ВШЭ. Из них отобраны 6 человек, имеющих наибольшее количество друзей из НИУ ВШЭ в социальной сети «ВКонтакте» (табл. 2).

Таблица 2 Структура «семян» выборки первого опроса

Респондент №	Пол	Количество друзей
1	Жен	264
2	Муж	399
3	Жен	287
4	Муж	473
5	Жен	390
6	Муж	207

³Данный участник был отобран несмотря на малое количество друзей, так как имел серьезную личную мотивацию.

Отобранным «семенам» на e-mail были разосланы инструкции с описанием содержания исследования и условиями сотрудничества. После заполнения анкеты респонденты должны были разослать аналогичные инструкции трем друзьям, обучающимся в НИУ ВШЭ. Для минимизации усилий «семян» к письму был приложен шаблон приглашения. Для повышения вероятности получения личной информации респондентам поясняли, что она будет использована только для выплаты вознаграждения. Кроме того, исследователи оставили собственные личные данные, по которым с ним можно было связаться в случае возникновения вопросов (ФИО, e-mail адреса, номера телефонов и ссылки на профиль «ВКонтакте»).

За первые 3 дня после рассылки инструкций «семенам» собрано 14 интервью, однако на этом рост выборки прекратился. Для его возобновления респондентам разослали два напоминания. В результате, онлайн-ВНР-1 достигла 72 респондентов (за исключением прерванных интервью и интервью, для которых не удалось установить рекрутера).

Онлайн-ВНР-2

Для участия в онлайн-ВНР-2 были приглашены 5 близких знакомых исследователей, обучающиеся в НИУ ВШЭ (табл. 3).

Таблица 3 Структура «семян» выборки второго опроса

Респондент №	Пол	Количество друзей
1	Жен	236
2	Муж	310
3	Жен	843
4	Муж	1232
5	Муж	303

Они также получили ссылки на анкеты и соответствующие инструкции. Шаблон для последующих респондентов не использовался. В результате без дополнительных напоминаний онлайн-ВНР-2 достигла 79 респондентов.

Результаты исследования

В ходе анализа полученных данных проверялось, соответствуют ли онлайн-ВНР-1 и -2 трем базовым допущениям: реципрокность связей между респондентами (1), способность респондентов правильно оценить размер своей эго-сети (2) и случайный отбор респондентов (3).

- 1 В данном исследовании реципрокность трактовалась как *взаимность связи*, основывающаяся на знакомстве респондента и рекрутера и выражающаяся в том, что они поддерживают дружеские отношения. Участникам задавался вопрос, в каких отношениях они состоят с человеком, который отправил им

- приглашение: «Этот человек является Вашим другом, знакомым или Вы не были знакомы с ним до участия в данном опросе?». При расчете распределений из анализа были исключены ответы «семян» и респондентов, которые не указали рекрутера. Незначительное расхождение полученных распределений позволяет предположить, что 1,5% незнакомых связей в сетевой выборке объясняется влиянием случайных факторов (ошибка при заполнении анкеты) и допущение № 1 обоснованно.
- 2 Для получения представления о том, насколько точно респонденты могут оценить размер своей эго-сети, использовался такой косвенный параметр, как оценка количества друзей в социальной сети «ВКонтакте». Помимо вопроса о том, какое количество друзей респонденты имеют в НИУ ВШЭ, им также задавался вопрос о числе приятелей в социальной сети. Если они могут корректно оценить последний параметр, оценка первого также окажется достаточно точной. Чтобы установить реальное количество друзей «ВКонтакте», авторы проанализировали профили респондентов.
 - 3 97,5% респондентов онлайн-ВНР-1 и 95,3% – онлайн-ВНР-2 имеют профиль в социальной сети «ВКонтакте», все они дали ответ о количестве друзей в этом сообществе. Сравнение оценки и реального количества друзей удалось установить для 19,4% и 13,7% опрошенных первого и второго опросов соответственно. Респонденты затрудняются корректно оценить количество друзей в социальной сети, хотя данное значение могло быть проверено ими в процессе заполнения анкеты. Допущенные ошибки (без учета возможного изменения количества друзей «ВКонтакте») в среднем составляют 12% для респондентов первого опроса и 18% для второго; в общей сложности 6 респондентов дали оценку, отличающуюся от реальной более чем на 20%.
 - 4 Предположение о случайном отборе адресата инструкции. При оценке этого допущения предполагалось, что на выбор рекрутера не должны оказывать влияние характеристики адресата приглашения, например, пол, возраст, курс, факультет обучения и условия проживания. Если это так, частота отбора членов одной группы должна совпадать с долей этой группы в генеральной совокупности. В противном случае на качестве выборки будет сказываться эффект гомофилии.

Согласно результатам онлайн-ВНР-1 и онлайн-ВНР-2, решения респондентов имеют отчетливое влияние эффекта гомофилии: например, мужчины оказались склонны приглашать к участию других мужчин (0,31 и 0,33 соответственно). Студенты, проживающие в общежитии, приглашали своих соседей (0,49 и 0,42). В онлайн-ВНР-1 высокий коэффициент гомофилии также проявили студенты, проживающие с родителями (0,45), и респонденты, проживающие в собственной квартире или снимающие жилье.

Таким образом, в рамках проведенных опросов не выполнилось одно из проверяемых базовых предположений ВНР о случайном отборе респондентов для рекрутирования. Также авторам не удалось достоверно установить, насколько в онлайн-ВНР выполняется условие о том, что любой респондент может попасть в выборку повторно. Возможно, кто-то получал инструкцию повторно, однако не стал еще раз заполнять анкету и приглашать друзей.

Оценка влияния личного знакомства «семян» с исследователями на качество полученной сетевой выборки

В целом сравнение результатов и динамики онлайн-ВНР-1 и -2 показывает, что набор респондентов проходит эффективнее, если «семенами» являются знакомые исследователя. Набор респондентов для первого опроса длился 2 месяца: первая онлайн-анкета заполнена 20 мая 2013 г., последняя – 14 июня 2013 г. Скорость набора респондентов для второго опроса увеличилась в 2 раза. Для онлайн-ВНР-2 не было необходимости использовать повторные приглашения.

Для анализа качества полученной выборки использован статистический пакет анализа данных, полученных при помощи ВНР – RDSAT версии 7.1.38.

Анализ проводился по всем контролируемым параметрам: пол, факультет, условия проживания, рабочая занятость респондента.

(1) Сравнение распределений респондентов первого и второго опросов по полу показало, что по этому признаку первая выборка оказалась более точной по отношению к генеральной совокупности (табл. 4).

Таблица 4 Оценка параметров генеральной совокупности по полу

Контрольный параметр		ГС	Ожидаемые значения, опрос 1	Ожидаемые значения, опрос 2
Пол респондента	Мужской	0,46	0,32	0,22
	Женский	0,54	0,69	0,78
Условия проживания	Живу в собственной квартире	0,07	0,06	0,04
	Живу в общежитии	0,35	0,41	0,58
	Живу с родителями	0,48	0,51	0,35
	Снимаю квартиру	0,09	0,03	0,03
Трудовая занятость	Да, работаю	0,34	0,346	0,134
	У меня нет оплачиваемой работы	0,67	0,654	0,866

(2) Распределение опрошенных онлайн-ВНР-1 и -2 по факультетам показало смещение в сторону экономического факультета (60% против 12% в генеральной совокупности), что оказало наибольший эффект в онлайн-ВНР-1.

(3) При оценке распределения параметров выборки по условиям проживания более точный результат получен по итогам онлайн-ВНР-1. Это может быть связано с тем, что большинство из знакомых исследователей проживают в общежитии, поэтому данные второго опроса сильно смещены в этом направлении – 58% против 35%. Параметр рабочей занятости респондента точнее представлен в онлайн-ВНР-1.

Приведенные данные демонстрируют, что качество полученных выборок не советуется главному критерию – способности репрезентировать генеральную совокупность. По одним параметрам лучшую оценку генеральной совокупности дает Web-ВНР-1, по другим – Web-ВНР-2, на основании этого результат можно считать случайным.

Ограничения и вопросы для обсуждения

Проведенное исследование — попытка реализации модели онлайн-ВНР, предложенной Хекаторном и его коллегами. Полученные результаты в целом говорят о том, что она не удалась или не привела к получению качественного результата. Данный вывод имеет ряд ограничений, связанных со строгостью проведенного исследования и возможностями оценки его результатов.

Проведенное нами исследование отличается от проектов Хекаторна по таким параметрам, как модель вознаграждения (у нас она не была отдельной), структура и количество «семян» (9 против 6 или 5). Хекаторн проводил эксперименты в студенческом кампусе, в то время как наш проект предполагал опрос респондентов, не проживающих на одной относительно ограниченной территории.

Кроме того, модели выборки первого и второго опросов отличаются не только по контрольному фактору (личное знакомство «семян» с исследователями), но и по таким важным параметрам, как время проведения опроса (весна и лето), количество «семян» (6 и 5).

Наконец, для оценки реализации базовых допущений нами использовались косвенные параметры, способные повлиять на точность и корректность итоговых результатов (например, проверка предположения о том, что респонденты могут точно оценить размер своей эго-сети делалась на основании их оценки количества друзей в социальной сети «ВКонтакте»).

С учетом обозначенных условий исследование показало, что приглашение респондентов, которые лично не знакомы с исследователем, работает неэффективно. Рост сети быстро прекращается и требует дополнительной стимуляции в виде напоминаний. Второй способ рекрутирования «семян» среди знакомых исследователя показал эффективность в динамике роста сети. Персональная ответственность первичного респондента перед исследователями, видимо, заставляла его мотивировать других участников сети. Можно предположить, что наиболее эффективным дизайном онлайн-ВНР является совмещение двух стратегий рекрутирования и увеличение количества «семян».

Литература

- 1 Девятко И. Ф. Инструментарий онлайн-исследований : попытка каталогизации // Онлайн-исследования в России 3.0 / под ред. А. В. Шашкина, И. Ф. Девятко, С. Г. Давыдова. М. : Типография КЕМ, 2012.
- 2 Мавлетова А. М. Проведение онлайн ВНР эксперимента : изучение мотивации гемблеров // Вестник Российского университета дружбы народов. Серия: Социология. 2010. № 4. С. 36–49.
- 3 Goel S., Salganik M. J., Raftery A. Assessing respondent-driven sampling // Proceedings of the National Academy of Sciences of the United States of America. 2010. Vol. 107, Nr 15. P. 6743-6747.
- 4 Heckathorn D. Extensions of respondent-driven sampling : analyzing continuous variables and controlling for differential recruitment author // Sociological Methodology. 2007. Vol. 37. P. 151–208.
- 5 Heckathorn D. Respondent-driven sampling : a new approach to the study of hidden populations // Social Problems. 1997. Vol. 44, Nr 2. P. 174–199.

- 6 Heckathorn D. Respondent-driven sampling II : deriving valid population estimates from chain-referall samples of hidden populations // Social Problems. 2002. Vol.49, Nr 1. P. 11–34.
- 7 Heckathorn D., Salgamik M. Sampling and estimation in hidden populations using respondent-driven sampling // Sociological Methodology. 2004. Vol. 34. P.193–239.
- 8 Heckathorn D., Wejnert C. Web-based network sampling efficiency and efficacy of respondent-driven sampling for online research // Sociological Methods & Research. 2008. Vol. 37. P. 105–134.
- 9 Wejnert C. An empirical test of respondent-driven sampling: point estimates, variance, degree measures, and out-of-equilibrium // Sociological Methodology. 009. Vol. 39. P. 73–166.